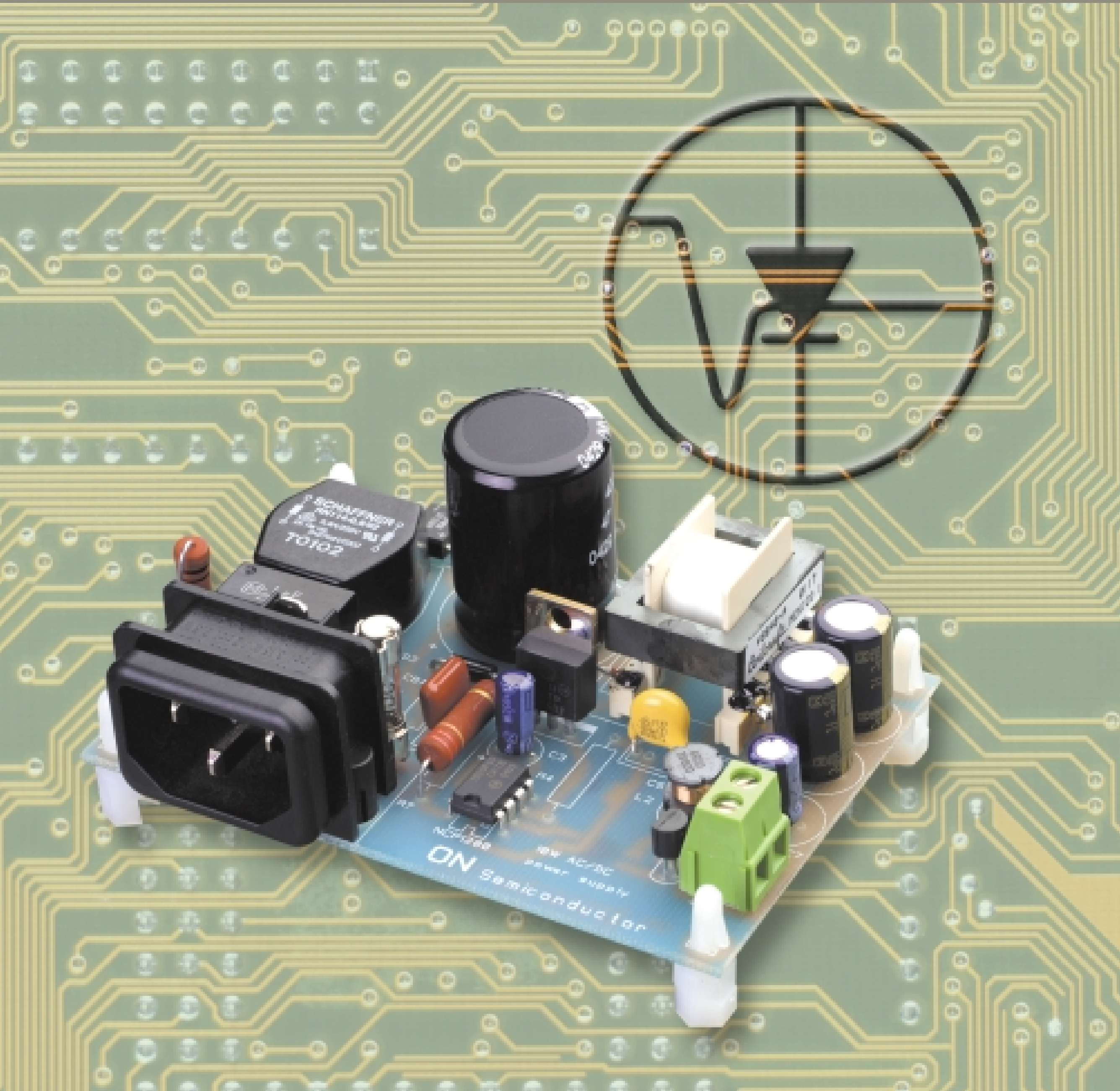
HB214/D
Rev. 2, Nov-2001

Rectifier Applications Handbook



Rectifier Applications Handbook

Reference Manual and Design Guide

HB214/D
Rev. 2, Nov–2001

ON Semiconductor™

© SCILLC, 2001
Previous Edition © 1997
"All Rights Reserved"

COVER ART
The NCP1200 10 W AC/DC Power Supply Demo Board featuring the MBRS360T3 and MUR160 Rectifiers. For more information, please see Appendix B on page 259.

Rectifier Applications

Acknowledgments

Technical Editor

William D. Roehr, Staff Consultant

Contributions

Samuel J. Anderson
Don Baker
Kim Gauen
David Hollander
Rex Ivins
Nick Lycoudes
Michael Pecht
William D. Roehr
William C. Roman
Howard Russell, Jr.
Aristide Tintikakis

Cho–Therm is a registered trademark of Chromerics, Inc.
Grafoil is a registered trademark of Union Carbide.
Kapton is a registered trademark of du Pont de Nemours & Co., Inc.
Kon Dux is a trademark of Aavid Thermal Technologies, Inc.
POWERTAP, SCANSWITCH, MEGAHERTZ, Surmetic, and SWITCHMODE are trademarks of Semiconductor Components Industries, LLC (SCILLC).
Sync–Nut is a trademark of ITW Shakeproof.
Thermalsil is a registered trademark and Thermalfilm is a trademark of Thermalloy, Inc.
Thermal Clad is a trademark of the Bergquist Company.

ON Semiconductor and ON are trademarks of Semiconductor Components Industries, LLC (SCILLC). SCILLC reserves the right to make changes without further notice to any products herein. SCILLC makes no warranty, representation or guarantee regarding the suitability of its products for any particular purpose, nor does SCILLC assume any liability arising out of the application or use of any product or circuit, and specifically disclaims any and all liability, including without limitation special, consequential or incidental damages. “Typical” parameters which may be provided in SCILLC data sheets and/or specifications can and do vary in different applications and actual performance may vary over time. All operating parameters, including “Typicals” must be validated for each customer application by customer’s technical experts. SCILLC does not convey any license under its patent rights nor the rights of others. SCILLC products are not designed, intended, or authorized for use as components in systems intended for surgical implant into the body, or other applications intended to support or sustain life, or for any other application in which the failure of the SCILLC product could create a situation where personal injury or death may occur. Should Buyer purchase or use SCILLC products for any such unintended or unauthorized application, Buyer shall indemnify and hold SCILLC and its officers, employees, subsidiaries, affiliates, and distributors harmless against all claims, costs, damages, and expenses, and reasonable attorney fees arising out of, directly or indirectly, any claim of personal injury or death associated with such unintended or unauthorized use, even if such claim alleges that SCILLC was negligent regarding the design or manufacture of the part. SCILLC is an Equal Opportunity/Affirmative Action Employer.

PUBLICATION ORDERING INFORMATION

Literature Fulfillment:
Literature Distribution Center for ON Semiconductor
P.O. Box 5163, Denver, Colorado 80217 USA
Phone: 303–675–2175 or 800–344–3860 Toll Free USA/Canada
Fax: 303–675–2176 or 800–344–3867 Toll Free USA/Canada
Email: ONlit@hibbertco.com

N. American Technical Support: 800–282–9855 Toll Free USA/Canada

JAPAN: ON Semiconductor, Japan Customer Focus Center
4–32–1 Nishi–Gotanda, Shinagawa–ku, Tokyo, Japan 141–0031
Phone: 81–3–5740–2700
Email: r14525@onsemi.com

ON Semiconductor Website: http://onsemi.com

For additional information, please contact your local Sales Representative.

Table of Contents

Chapter 6. Polyphase Rectifier Circuits

Chapter 7. Rectifier Filter Systems

Chapter 8. Rectifier Voltage Multiplier Circuits

Chapter 9. Transient Protection of Rectifier Diodes

Chapter 10. Basic Diode Functions in Power Electronics

Chapter 11. Power Electronic Circuits

Chapter 1

Power Rectifier Device Physics and Electrical Characteristics

Power Rectifier Device Physics and Electrical Characteristics

The physics of power rectifiers is different in many aspects from that of their low–power counterparts. The reason for this is that power rectifiers must process power with a minimum power loss with operating conditions often covering a wide range of blocking voltages, current densities, and frequencies.

The ability to handle forward currents from 1 A to 600 A. block voltages up to 100 V, and exhibit excellent switching behavior at frequencies up to 1 MHz are the unique performance features of today's silicon Schottky rectifier technology Silicon bipolar junction rectifiers can block voltages up to 1500 V and have current capability up to 190 A. However, in switching applications silicon bipolar junction rectifiers require minority barrier life–time reduction to perform the switching behavior at the speeds necessary for switching operations These devices have higher conduction losses than slower standard and fast–recovery bipolar junction rectifiers and are selected when switching loss dominates.

Circuit designers need low–loss, power rectifiers both in conduction mode and dynamic mode operations, especially in high–frequency applications In this domain silicon–based technology is rapidly approaching its performance limit and the current industry demand for low–loss diodes is driving the quest for new semiconductor materials such as gallium arsenide Schottky devices for improved performance.

To provide the necessary background for selecting and using power rectifiers, this first chapter presents a review of basic semiconductor physics covering energy bands, surface properties. doping. conductivity, p–n junction and m–s junction theory. From this, the diode equations for ideal bipolar p–n junctions and majority carrier m–s junctions are developed. Next, properties of real diodes are shown; finally, possible tradeoffs in performance characteristics are summarized.

Formation of Energy Bands

Niels Bohr provided a model of the hydrogen atom in which a single electron orbits the nucleus and is held in orbit due to the attraction of a proton in the nucleus of the atom. Bohr added a rule of quantization to postulate that a discrete set of orbits exists from the continuum of possible orbits and this explains the discrete energy levels needed to describe the line spectrum of atomic hydrogen.

If we visualize the electron in a stable orbit of radius r about a proton of the hydrogen atom, we can equate the electrostatic force between the charges to the centripetal force, as shown in Figure 1.

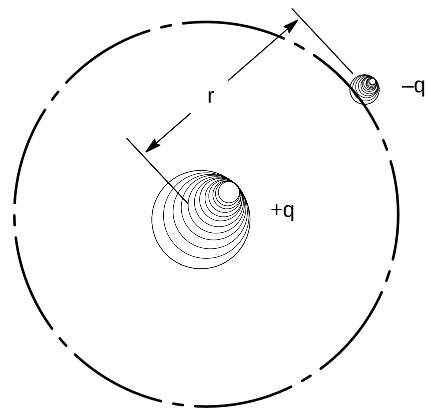


Figure 1. Bohr Hydrogen Atom Model

$$\frac{-q^2}{4,,\epsilon_o r^2} = \frac{mv^2}{r} \quad (1.1)$$

where:

m = mass of the electron

v = velocity of the electron

Since the electron is confined to the vicinity of the nucleus bound by a coulombic force, Bohr demonstrated that the possible electron energies are quantized. Hence, the energy states of the electron in the hydrogen atom are given by:

$$E_n = \frac{-m_o q^4}{8h^2\epsilon^2} \cdot \frac{1}{n^2} \quad (1.2)$$

where:

m_o is the electron mass and h is Planck's constant.

Materials used in semiconductor technology can be described in terms of energy levels that electrons in an atom can occupy. For the purpose of this chapter, we will focus on energy levels theory. For a single crystal atom, the energy level perspective can be considered in terms of an orbital plan and an energy well plan. Both of these are shown in figure 2 for the silicon atom. Since intrinsic silicon has a valence of four and forms covalent bonds with four neighboring silicon atoms when more than one atom is present, energy levels combine as the atoms are joined.

As the silicon crystal is formed, the interatomic spacing is reduced and electron conduction bands arc formed by bringing together isolated silicon atoms. Figure 3(a) shows the formation of energy bands as the atomic separation is reduced during crystal formation. At absolute temperature 0°K every state in the valence band will be filled, while the conduction band will be completely empty of electrons. The arrangement of completely filled and empty energy bands has an important effect on the electrical conductivity

of a solid. Every solid has its own characteristic energy band structure. The primary difference between an intrinsic semiconductor, an insulator, and a good conductor is the size of the energy bandgap. Silicon has a bandgap of 1.1 eV, compared to 8 eV for an oxide insulator. The relatively small bandgap of silicon allows excitation of electrons from the lower valence band shown in Figure 3(b) to the upper conduction band by a reasonable amount of thermal or optical excitation energy. This is why silicon is a semiconductor. In metals the valence and conduction bands overlap so that electrons can move freely, resulting in high electrical conductivity.

The energy band diagram of intrinsic silicon is shown in Figure 4. The intrinsic silicon valence band is completely filled with valence electrons at absolute zero and no current can flow. But at any temperature above: absolute zero, a small number of electrons possess enough thermal energy to be excited to the conduction band. For every electron excited to the conduction band, a hole is created in the valence band. The intrinsic concentration of holes and electrons is relatively small for silicon at room temperature.

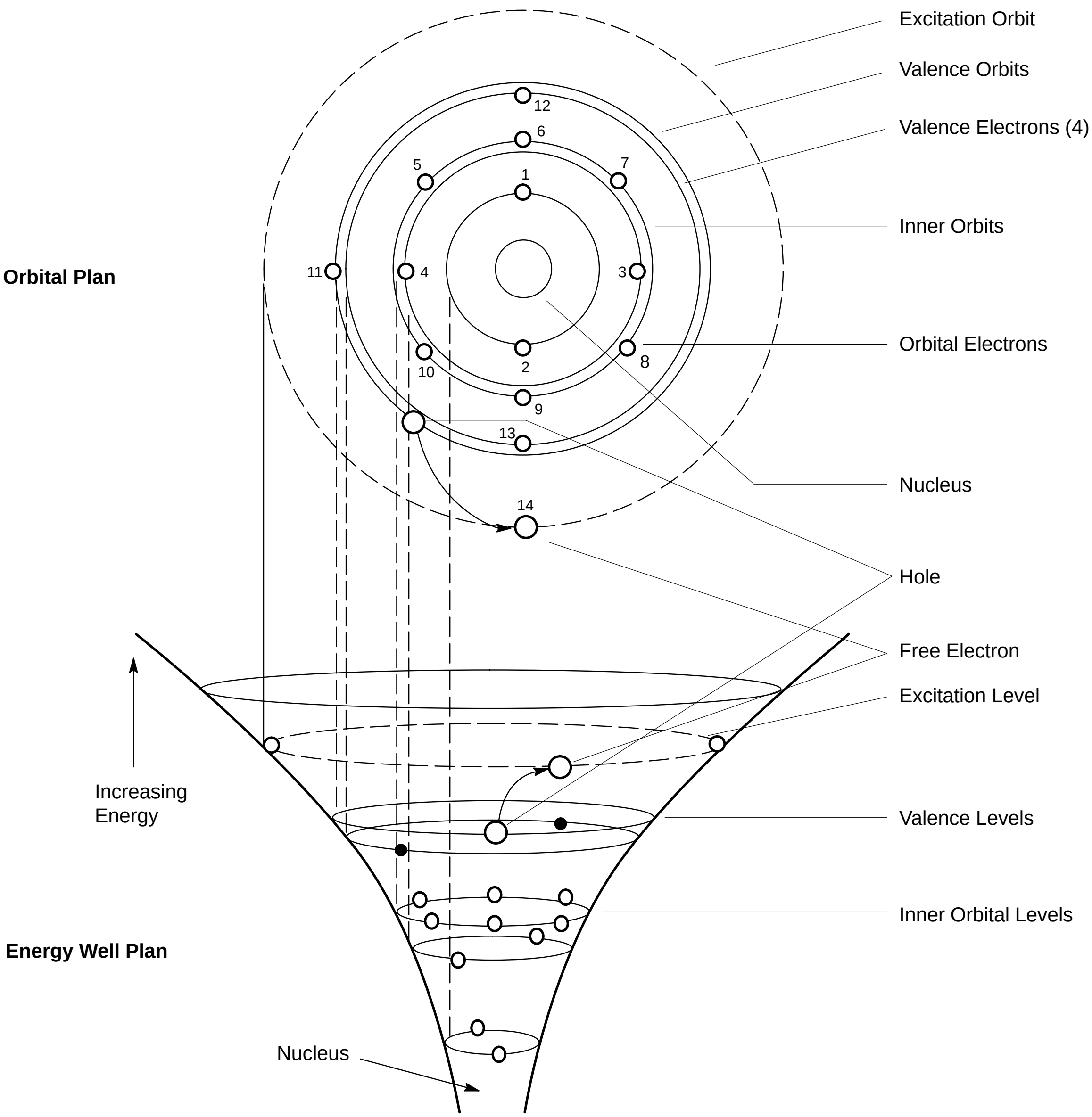


Figure 2. Orbital and Energy Well View of Electrons in Silicon

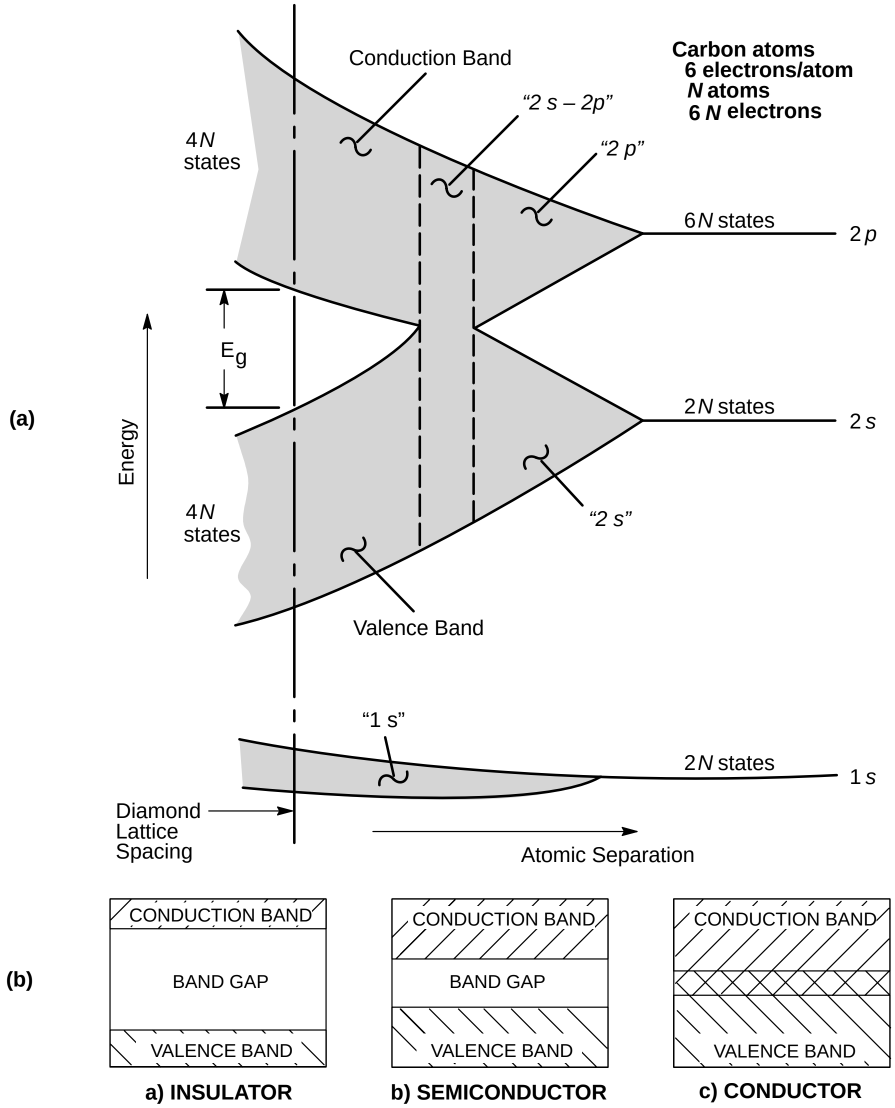


Figure 3. (a) Energy Band Versus Atomic Separation Relationship; (b) Energy Band Gap of Insulators, Semiconductors, and Conductors

However, the intrinsic concentration doubles every 10°C and is defined by Equation 1.3.

In intrinsic silicon, the number of conduction electrons, n, equals the number of holes, p, or

$$n = p = n_i = 3.9 \times 10^{16} T^{3/2} \exp\left(\frac{-E_G}{kT}\right) \text{cm}^{-3} \quad (1.3)$$

where:

n = electron concentration, cm^{-3}
p = hole concentration, cm^{-3}
n_i = the intrinsic carrier concentration /cm^3
T = temperature, °K
E_G = bandgap, eV
k = Boltzman's constant
at T = 00°

At 25°C n_i for intrinsic silicon is:

n_i = 1.5 X 10^{10} carriers/cm^3

whereas:

n_i = 1.8 X 10^6 carriers/cm^3 for GaAs.



Figure 4. Energy Band Diagram for Intrinsic Silicon

Doping of Semiconductors

Semiconductor materials have no practical application in their intrinsic form. To obtain useful semiconductor devices, controlled amounts of dopant atoms arc introduced into the crystal. When dopant atoms are added, energy levels are created near the conduction band for n–type semiconductors and near the valence band for p–type semiconductors. The band diagrams for n–type and p–type semiconductors are shown in Figure 5.

The addition of dopant atoms increases either the electron or hole concentration of the host silicon crystal. Common silicon dopants are listed in Table 1.

The result of doping is a change in the electrical conductivity of the crystal. If pentavalent elements such as antimony, phosphorus or arsenic are introduced into the silicon crystal, atoms will be displaced by the impurity atoms as illustrated in Figure 6(a). Four electrons of the impurity atom form covalent bonds with four surrounding atoms and the fifth electron is free to act as a carrier. The pentavalent impurity is called a donor because it gives up electrons and the material is said to be n–type because it is negatively charged with free electrons and donates electrons to the conduction band.

Table 1. Common Silicon Dopants

Electron–Increasing Dopants (Donors)	Hole–Increasing Dopants (Acceptors)
P	B
As *Column V*	Ga *Column III*
Sb *Elements*	In *Elements* Al

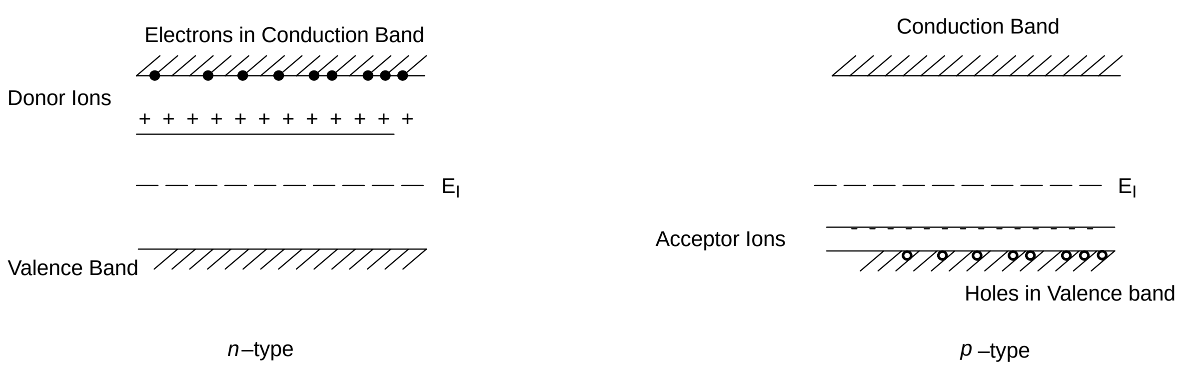
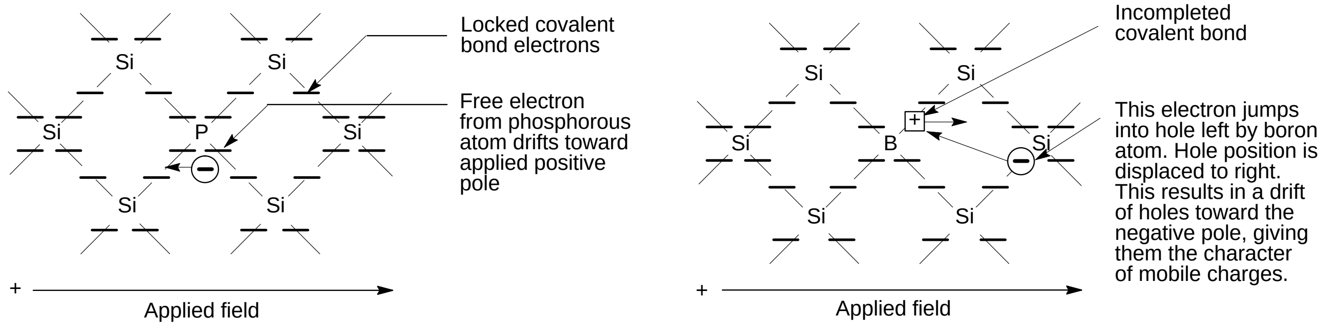


Figure 5. Band Diagram for *n*–type and *p*–type Semiconductors



a) n–type Silicon **b) p–type Silicon**

Figure 6. Impurity Atoms in (a) *n*–type Silicon; (b) *p*–type Silicon

If a trivalent impurity atom such as boron, gallium, indium, or aluminum is added to the semiconductor, the electrons surrounding the impurity atom cannot form a stable set of locked covalent bonds The tri–electron structure of the impurity atom causes one incomplete bond as shown in Figure 6(b). The result is a positively charged material. The trivalent impurity is called an acceptor impurity because it accepts electrons from the valence band and forms holes in the valence band.

When donors are added to silicon, there is a larger number of conduction electrons or holes, respectively. It can be shown mathematically that

$$n_i^2 = n \cdot p = N_D^+ \cdot N_A^- = 2 \times 10^{20} / \text{cm}^6 \text{ at } 26^\circ\text{C} \qquad (1.4)$$

where:

N_D is the doping concentration per cubic centimeter for donors and N_A is the acceptor doping concentration per cubic centimeter.

Therefore, if n or p is known, the other carrier concentration can be readily found.

The Electrical Resistivity and Conductivity of Silicon and Gallium Arsenide

Conduction of electricity through bulk silicon or gallium arsenide is by majority carriers Resistivity is the inverse of conductivity For n–type material it is given by:

$$\sigma_n = nq\mu_n \qquad (1.5)$$

where:

σ_n = conductivity $(\Omega\text{–cm})^{-1}$
n = electron concentration, cm^{-3}
q = electron charge
μ_n = electron mobility, $\text{cm}^2/\text{V–S}$

$$\text{Pn is } 1/\sigma_n = \frac{1}{nq\mu_n} = \text{ resistivity} \qquad (1.6)$$

The mobility described above is the proportionality constant between the applied electric field (ε) and the resultant carrier drift velocity (V_d) for bulk carrier transport where:

$$\mu_n = \frac{V_d}{\varepsilon} \text{(obeys Ohm's law)} \qquad (1.7)$$

Figure 7 shows bulk carrier mobility as a function of doping concentration for both electrons and holes.

Drift velocity is determined by the mean free time (T_f) between collisions within the crystal lattice, this being the time they are accelerated in the direction of the field. Using Newton’s law of motion, the drift velocity is given by:

$$V_d = \frac{q\epsilon T_f}{2m*} \qquad (1.8)$$

where:

m* = conductivity effective mass
ε = applied electric field
T_f = mean tune between collisions in the lattice

since:

$$\mu_n = \frac{V_d}{\epsilon} = \frac{qT_f}{2m} \qquad (1.9)$$

In uniform extrinsic crystals where the carrier concentration is predominantly of one type, the electrical conductivity σ_n, is determined by the acceleration of carriers due to the applied electrical field and those processes which arrest carrier motion through the lattice.

$$\sigma_n = qn\mu = qn\frac{V_d}{\epsilon} = q^2n\frac{t}{m*} \qquad (1.10)$$

m* = 0.98 for silicon
m* = 0.067 for gallium arsenide

Since the effective: mass of gallium arsenide is much lower than that of silicon, the force on an electron in an electric field accelerates electrons in gallium arsenide much more quickly than electrons in silicon. For low field conduction ($<10^3$ V/cm) the scattering processes for gallium arsenide and silicon are similar in magnitude, so the mean drift velocity for low–energy carriers is much higher, resulting in a low–field electron mobility of 8500 cm–V–s^{-1} at 300°K for gallium arsenide and 1500 cm–V–s^{-1} at 300°K for silicon.

The effective mass of the atom determines the mobility and hence conductivity of the semiconductor material. For power rectifiers this property offers the potential for faster

devices with lower on–resistance when gallium arsenide is selected over silicon.

By taking into account the fact that mobility changes with doping concentration,one can reconstruct Irvin's curves which plot resistivity as a function of doping concentration, as shown in Figure 8 for silicon and Figure 9 for gallium arsenide.

From Figure 8 if a concentration of 10^{16} cm^{-3} is selected, this equates to a resistivity of 0.5 Ω–cm for silicon. This is the typical resistivity tax–get used for a 40 V power Schottky. From Figure 9, for the same concentration of 10^{16} cm^{-3}, a resistivity of 0.05 Ω–cm is obtained with gallium arsenide as the semiconductor material for again a 40 V power Schottky device. This demonstrates the feature of low on–resistance potential with gallium arsenide over silicon but other problems pertaining to surface properties (which will be discussed later in this chapter) dominate the forward voltage drop in gallium arsenide Schottky devices.

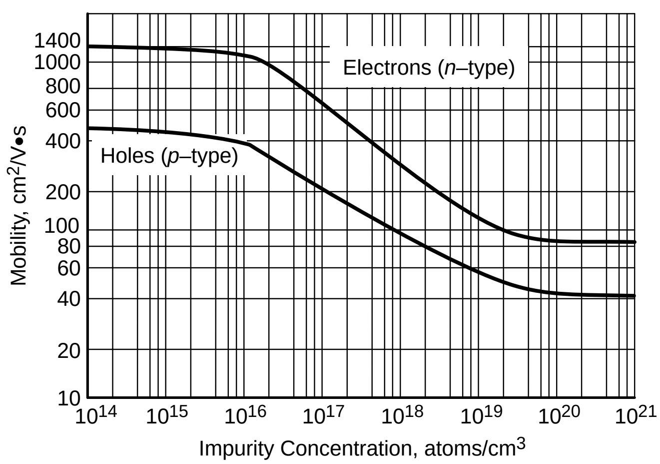


Figure 7. Bulk Mobility Versus Doping for Silicon

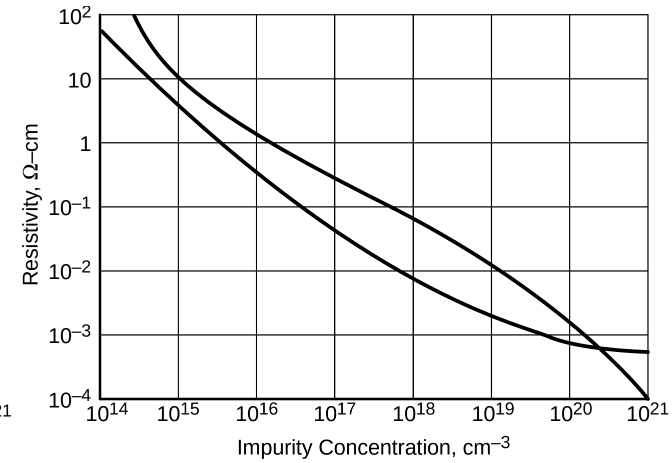


Figure 8. Resistivity as a Function of Doping Concentration (Irvin's Curve) for Silicon [1]

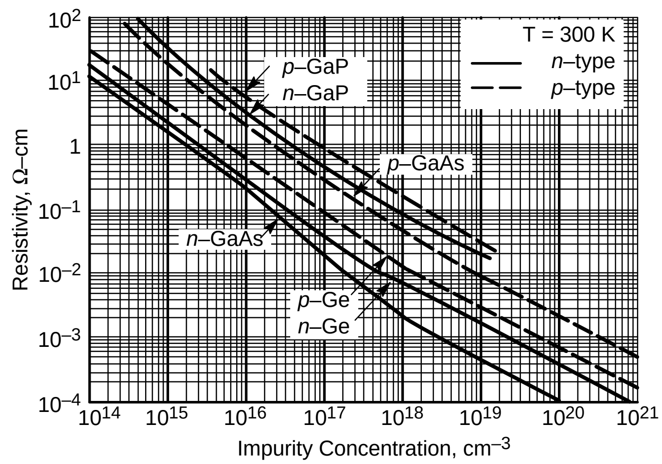


Figure 9. Resistivity Versus Impurity Concentration for Ge, GaAs, and GaP at 300°K [2]

The Fermi Level Concept

The fermi level is used to describe the energy level of the electronic state which one electron has a probability of 0.5 of occupying. However, the fermi level also describes the number of electrons and holes in the semiconductor.

With reference to the energy band diagram of Figure 10

$$n = n_i \exp\frac{E_f - E_i}{kT} \qquad n - \text{type} \qquad (1.11)$$

$$p = n_i \exp\frac{E_i - E_f}{kT} \qquad p - \text{type} \qquad (1.12)$$

for intrinsic since $E_f = E_i$

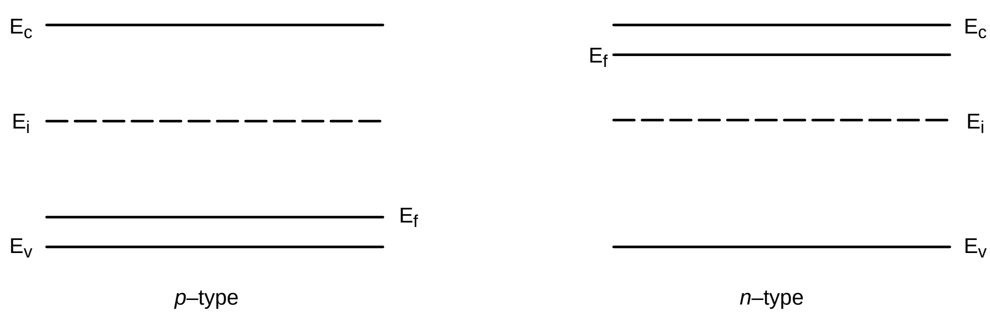
$$p = n = n_i \qquad (1.13)$$

In other words, for the intrinsic material, the fermi level is at the midpoint of the band gap.

Surface Properties of Semiconductors

The atoms in a crystal of silicon or gallium arsenide are arranged in a well–ordered structure so that the net force acting on each atom is zero. In the bulk crystallographic structure of silicon, each atom is surrounded by four others in a tetrahedral configuration. except for imperfections such as vacancies, substitutional or interstitial impurities, or dislocations, this configuration extends throughout the single crystal. The situation at the surface of a crystal of silicon or gallium arsenide is different from that prevailing in the bulk in many important ways. First of all, the absence of neighboring atoms causes the equilibrium positions adopted by the surface atoms to be different from the bulk. Secondly, the chemical nature of the surface tends not to be atomically clean (i.e., the atomic composition of the surface is the same as the bulk). The surfaces of semiconductors are complex regions where the chemistry and crystallography is quite different from the bulk Not surprisingly, the distribution of electrons and associated energy levels also differs at the surface. from those pertaining to the bulk.

Since these surfaces are critical in the formation of oxide–semiconductor interfaces and metal–semiconductor interfaces, it is important understand them to the fullest possible extent. Consider a semiconductor material like silicon freshly cleaved in high vacuum to expose an ultraclean surface. From a symmetry viewpoint, two of the four covalent bonds which hold the silicon lattice together would be broken at the crystal surface in contact with the vacuum. These broken bonds are missing two electrons per atom, and hence, about. 10^{15} atoms/cm^3 are in a position to attract negative charges onto the crystal surface and give rise to energy states in the forbidden energy gap. In practice, the silicon surface atoms when exposed to air quickly form a native oxide layer about 40Â thick. This satisfies the silicon "dangling" bonds at the surface and reduces surface states under clean conditions; to 5 x 10^{10}/cm^2 or less. Figure 11(a) shows the energy band diagram for the interface between a p–type silicon crystal and a vacuum, assuming a perfect crystal in a perfect vacuum ignoring the effects of broken bonds at the surface This may be referred to as the flat–band case, but obviously does not occur in practice: The real situation is shown in Figure 11(b). The SiO_2 layer contains positive charges that result in an n–shift in the semiconductor, i.e., an n–type semiconductor behaves more n–type, whereas p–type material behaves less p–type.



E_c = Energy level of conduction band
E_v = Energy level of valence band
E_f = Fermi energy level
E_i = Intrinsic energy level

Figure 10. Energy Band Diagram for n–type and p–type with Fermi Levels Shown

Rectifier Applications

With respect to the clean surfaces of semiconductors and metals, surface states also significantly impact device properties. The deposition of a metal onto a semiconductor usually gives rise to a barrier, if the work function fm of the metal is greater than the work function of the semiconductor. For the case of a metal to n–type semiconductor, the barrier height (f_B) is equal to the separation of the fermi level (E_f) and the conduction band minimum E_c at the semiconductor metal interface. Schottky proposed that f_B be given by the difference between the metal work function and the semiconductor electron affinity (X_S). With measured barrier height for metals on n–type GaAs usually in the range of 0.7 to 0.9 eV, the Schottky model fails to explain the observed insensitivity of the barrier height to metal work function in gallium arsenide. The basic explanation for this is thought to be surface states and interface states causing the fermi level to be pinned at a certain level between the conduction and valence band at the semiconductor surface. Unpinning of the GaAs fermi level with heavily doped silicon overlayers has been reported [1] but significant work in this area remains to be done to gain complete understanding and allow realization of improved on–conduction performance of gallium arsenide Schottky technology.

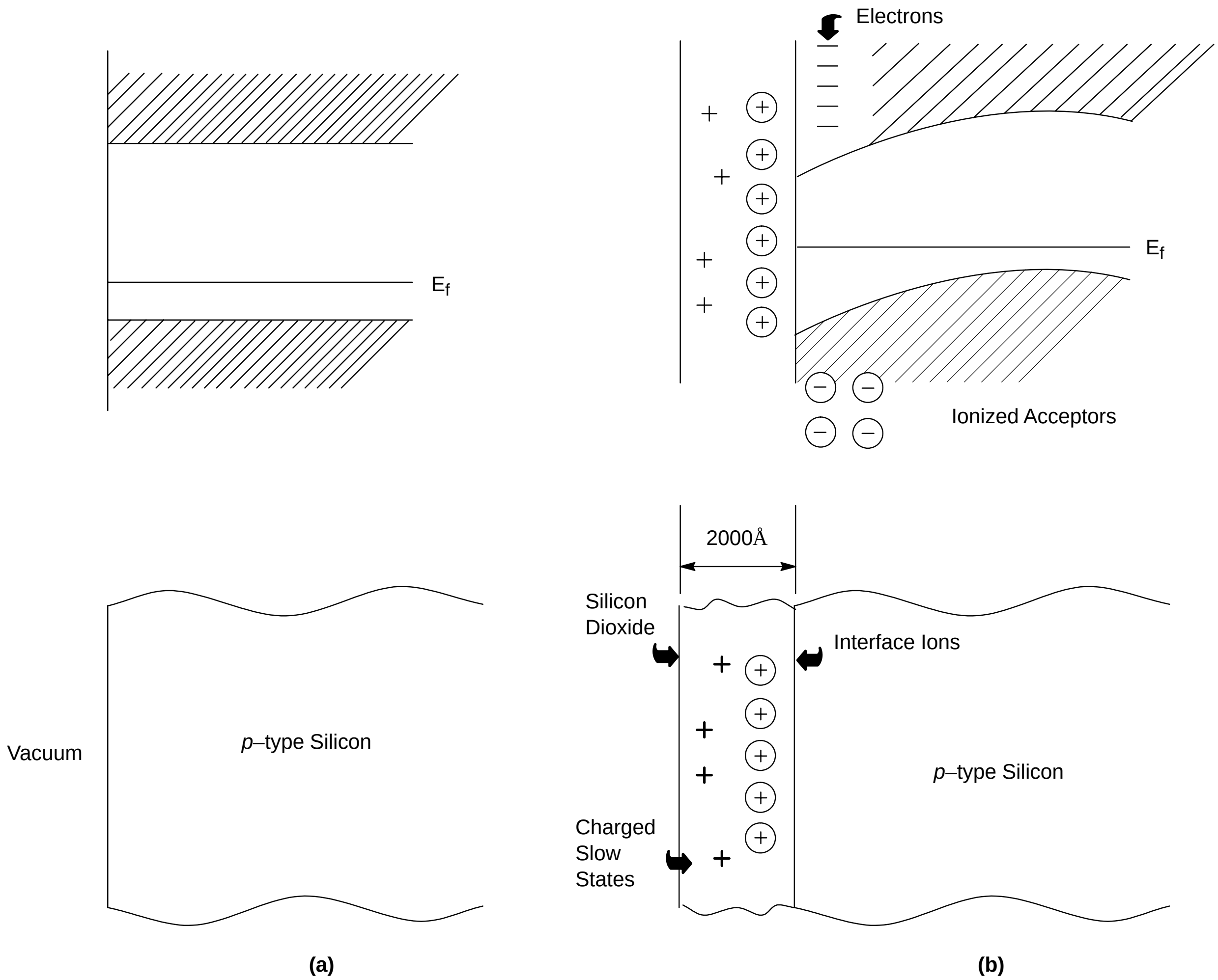


Figure 11. (a) Ideal Surface; (b) Band Rending Due to Energy States in the Forbidden Energy Gap

Semiconductor Junctions

pn Junctions

The rectifying properties of a diode can be understood most easily from its band diagram. Prior to junction formation, the band diagram of separated n–type and p–type silicon is shown in Figure 12. From Equation 1.6:

$$P = n_i \exp(Ei - E_F)/kT \qquad (1.14a)$$

From Equation 1.5:

$$P = n_i \exp(E_F - Ei)/kT \qquad (1.14b)$$

Hence,

$$kT \ln (p/n_i) = Ei - E_F \quad \text{and}$$

$$kT \ln (n/n_i) = E_F - Ei \qquad (1.15)$$

and therefore:

$$Ei - E_f = dT \ln (p/n_i) \quad \text{and}$$

$$E_F - Ei = kT \ln (n/n_i) \qquad (1.16)$$

When the two separated regions are joined, energy bands bend to give a flat fermi level when an external voltage is applied as shown in Figure 13.

$$V_{bi} = (Ei - Efown20F)_p \text{ side} + (E_F - Ei)n \text{ side} \qquad (1.17)$$

$$V_{bi} = kT (p/n_i) + kT \ln (n/n_i) \qquad (1.18)$$

$$V_{bi} = kT \ln [pn/n_i^2] \qquad (1.19)$$

Since p = N_A and n = N_D then:

$$V_{bi} = kT \ln [N_A N_D/n_i^2] \qquad (1.20)$$

Once the equilibrium point is reached after junction formation, the electrons on the n–type side cannot flow across to the p–type side because of the energy barrier due to the “built–in” contact potential similarly, holes on the p–type side cannot flow across to the n–type side.

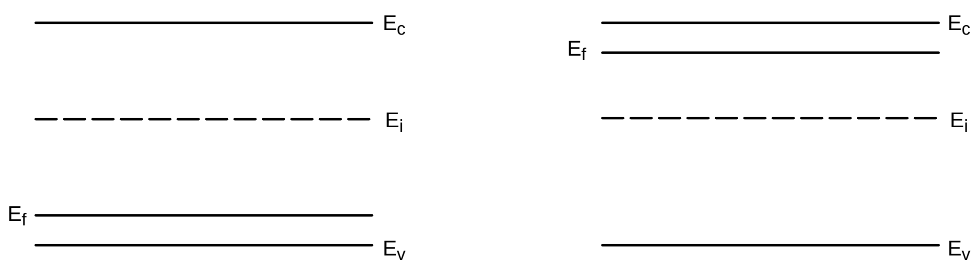


Figure 12. Band Diagram Prior to Junction Formation

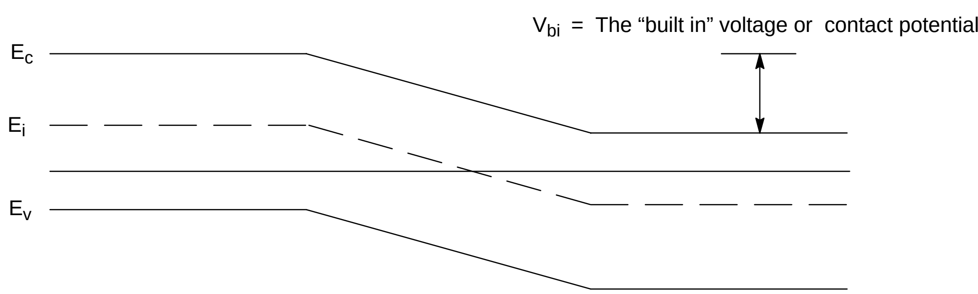


Figure 13. Band Diagram of Ideal Pit Junction

Forward Bias

When an external voltage is applied to the junction, the fermi level is offset at the terminals by the amount of the applied voltage TIP a positive voltage is applied to the p–side of the junction, the pn junction is forward biased. This applied forward bias reduces the "built–in" barrier height and carriers' flow across the barrier, as shown in Figure 14.

In effect. the built–in potential is now reduced, causing a diffusion current of majority carriers to flow across the depletion region. Only holes with energies greater than the "built–in" voltage qV_{bi} can diffuse into the n–region. In addition, holes in the n–region near the depletion edge drift under the influence of the electric field in a direction opposite to the hole diffusion current, constituting a negative hole amount.

Total hole current

$$J_p = J_p/\text{drift} + J_p/\text{diffusion} \qquad (1.21)$$

The carrier flow of holes from the p–side to n–side shows as an increase in minority carrier concentration on the n–side of the junction. The excess minority carriers recombine outside of the depletion layer, falling to the equilibrium value within one diffusion length L.

Similarly, total electron current

$$J_n = J_n/\text{drift} + J_n/\text{diffusion} \qquad (1.22)$$

The net effect of forward bias is a large diffusion current component, while the drift component remains fixed near the thermal equilibrium value.

Reverse Bias

When the voltage applied to the p–type side is negative, the junction is reverse–biased. The applied voltage adds directly to the "built–in" voltage and causes the fermi level of the p–side to be displaced by an amount qV_A. The net result is that minority carriers are drawn toward the junction, and the hole diffusion current is reduced to less than its thermal equilibrium value. The reverse bias current is extremely small because the source of current flow, the minority carrier concentration, is small.

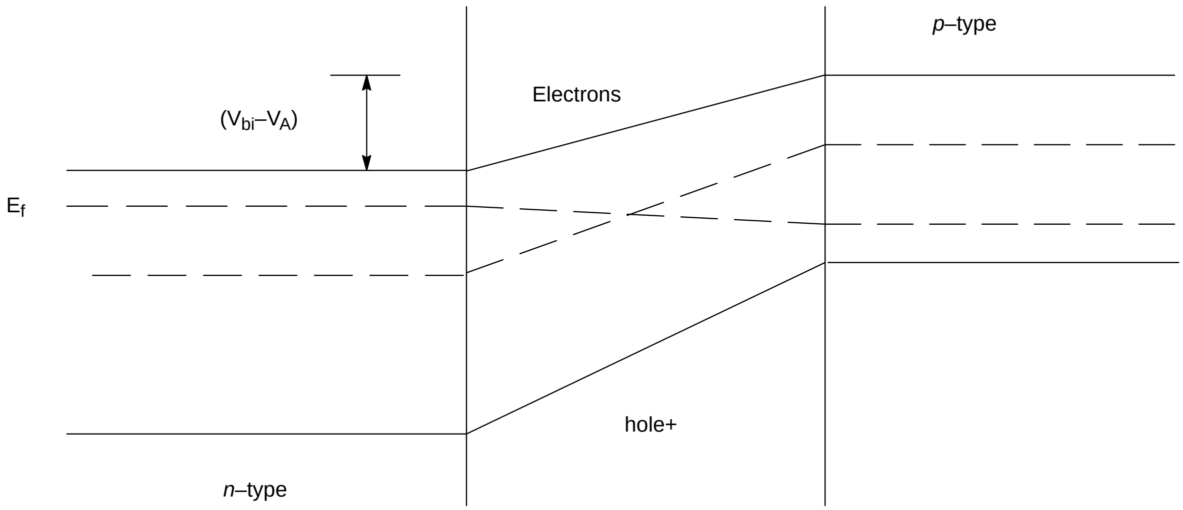


Figure 14. Band Bending Due to Forward Bias Reduces "Built–in" Potential

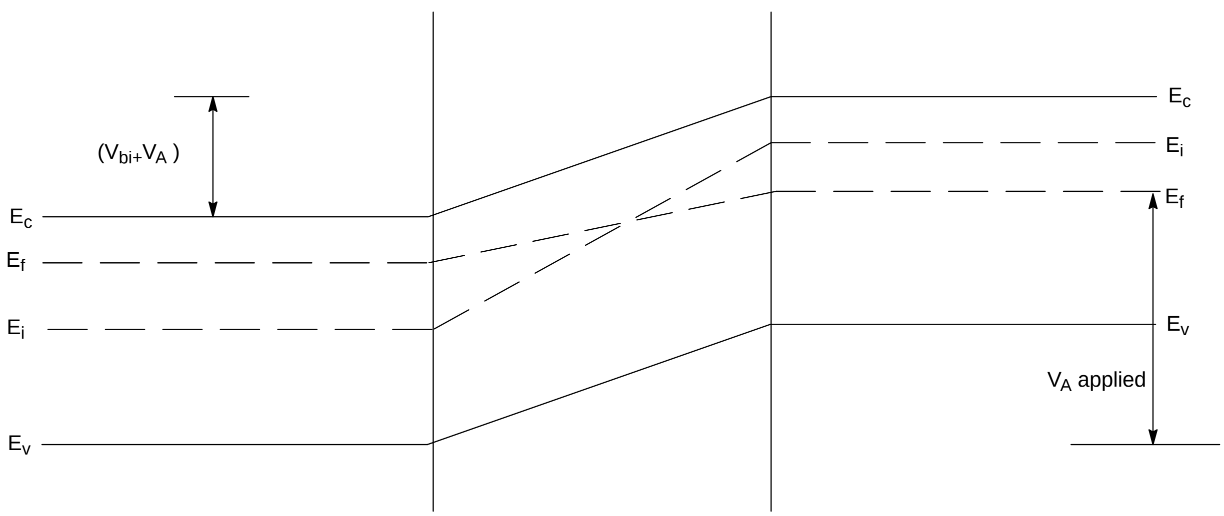


Figure 15. Band Bending Due to Reverse Bias

Metal–Semiconductor Junctions

Metal–semiconductor rectifying contacts are formed when the metal barrier is brought into contact with the lightly doped semiconductor material, forming an energy barrier between the metal and the semiconductor The barrier height of the metal–to–semiconductor junction depends on the semiconductor material and its surface properties as explained earlier in this chapter. The energy band diagram of a rectifying metal–semiconductor contact is shown in Figure 16.

According to the Schottky model, the barrier height can be determined by the difference between the metal work function (f_m) and the electron affinity X_S of the semiconductor. However, in most practical situations a thin insulating oxide film exists at the metal–semiconductor interface Such an insulation film, often referred to as an interfacial layer, alters the properties of the semiconductor surface from the bulk semiconductor (see section on surface properties). The existence of this interfacial layer enhances the density of surface states near the surface and as a result the barrier height does not follow the Schottky model of:

$$\phi_b = \phi_m - X_S \qquad (1.23)$$

Where ϕ_b = barrier height, ϕ_m = work function of metal, X_s = electron affinity of the semiconductor.

The barrier height consists of four components.

$$\phi_b = \phi_m - X_S - \Delta_{SS} - \Delta_\phi \qquad (1.24)$$

Where Δ_{SS} = contribution due to surface states and A.1. is image force lowering [3].

In practical processing of Schottky contacts, emphasis is placed on making the metal–semiconductor contact close to the ideal situation where the semiconductor is sensitive to the metal work function. For silicon Schottky technology this is accomplished by obtaining an intimate metal–to–semiconductor contact through the process of forming a thin silicide layer under the top metal contact. Table 2 shows the possible ranges of silicides that have been achieved to gain control of the barrier height on silicon–based Schottky rectifiers. As yet, control of the barrier height on gallium arsenide has proven elusive due to the pinning of the fermi level.

Table 2. Schottky Barrier Heights (Φ_B) of Various Silicides On *n*–Type Silicon

Disilicides	Φ_B(eV)	Other Silicides	Φ_B(eV)
$TiSi_2$	0.6	HfSi	0.53
VSi_2	0.65	MnSi	0.76
$CrSi_2$	0.57	CoSi	0.68
$ZrSi_2$	0.55	NiSi	0.7
$NbSi_2$	–	Ni_2Si	0.7
$MoSi_2$	0.55	RhSi	0.74
$HfSi_2$	–	Pd_2Si	0.74
$TaSi_2$	0.59	Pt_2Si	0.78
WSi_2	0.65	PtSi	0.87
$FeSi_2$	–	IrSi	0.93
$CoSi_2$	0.64	Ir_2Si_3	0.85
$NiSi_2$	0.7	$IrSi_3$	0.94

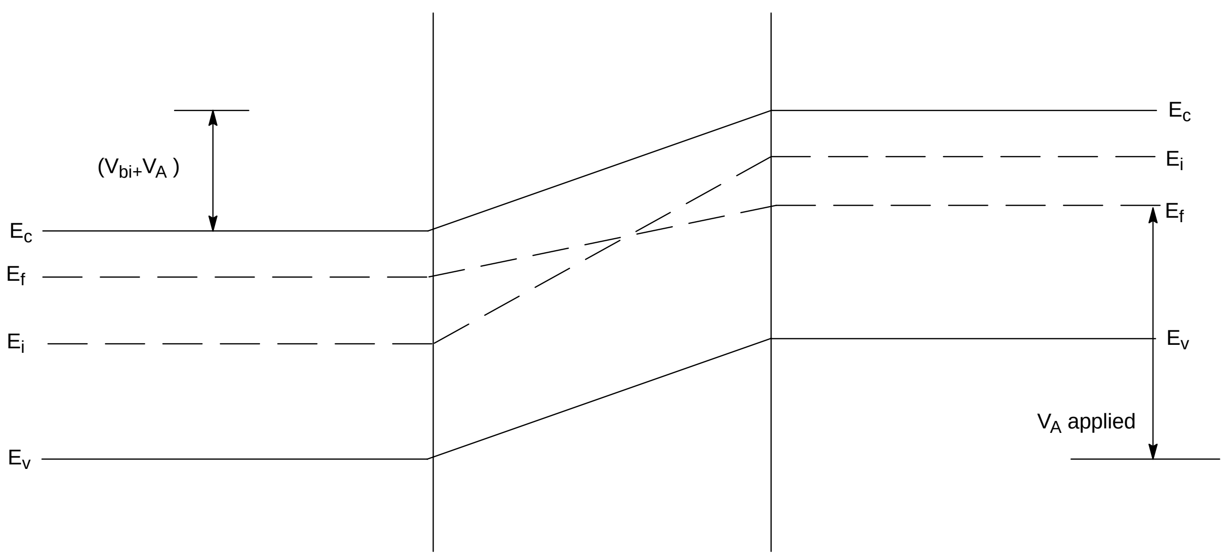


Figure 16. Schottky Energy Band Diagram (No–bias)

Since the work function of the metal and the electron affinity of the semiconductor are properties of the two materials selected to form the Schottky contact, the barrier height to electrons remains constant and does not vary with applied voltage. Under no bias, electrons at the edge of' the conduction band in the bulk semiconductor see a potential barrier qV_{bi} (Figure 17(a)). The number of electrons with energies greater than Ec + qV_{bi} traveling into the metal constitute a semiconductor–to–metal current density component Js–m. Since under no bias the Schottky diode is in equilibrium, the number of electrons transversing the metal–semiconductor junction from metal to semiconductor is:

$$J_{S-M} = -k_1 N_D e^{-qVbi/kT} \quad (1.25)$$

$$J_{M-S} = -k_1 N_C e^{-\phi_D/kT} \quad (1.26)$$

where:

N_c is the effective density of the conduction band states

N_A is the hulk majority carrier concentration from $n \equiv N_D$.

Forward Bias (V_A>0) and Reverse Bias (V_A<0)

The applied voltage changes the potential barrier for electrons. In the semiconductor but not for electrons in the metal. Therefore, $J_{S–M}$ remains constant whereas $J_{M–S}$ increases with increasing V_A. With V_A applied, the barrier for electrons in the semiconductor reduces to $q(V_{bi},–V_A)$ and becomes (see Figure 17(b)):

$$J_{S-M} = k_1 N_D e^{-q(Vbi-V_A)/kT} \quad (1.27)$$

The total current is $J_{S-M} + J_{M-S}$ which is

$$J = k_1 N_D e^{-qVbi/kT} e^{qV_A/kT} - k_1 N_D e^{-qVbi/kT} \quad (1.28)$$

$$J = k_1 N_D e^{-qVbi/kT} [e^{qV_A/kT} - 1] \quad (1.29)$$

Since:

$$J_{M-S} = k_1 N_C e^{-\phi D/kT} = -k_1 N_D e^{-qVbi/kT}$$

then:

$$N_D e^{-qVbi/kT} = N_C e^{-\phi D/kT}$$

Since Φ_B and V_{bi} do not change with V_A then:

$$J = k_1 N_C e^{-\phi D/kT} [e^{qV_A/kT} - 1] \quad (1.30)$$

$$J = J_0 [e^{qV_A/kT} - 1] \quad (1.31)$$

which is the form of the general diode equation.

According to thermionic emission theory, to describe the charge transport of electrons over a potential barrier the general form of the diode equation can be rewritten as

$$J = AT^2 e^{-(q\phi D/kT)} (e^{qV_A/kT} - 1) \quad (1.32)$$

where

J	=	Current flow across the Schottky barrier interface
A	=	Richardson's constant
T	=	absolute temperature
q	=	electron charge
k	=	Boltzman's constant
ϕ_B	=	barrier height
V	=	applied voltage

From Figure 17 the potential barrier for electrons in the semiconductor increases and results in a significant decrease in the number of electrons that can get to the metal. However, electrons in the metal see the same barrier as before f_B;f_B does not change with reverse bias. Therefore, the reverse current becomes

$$J_R = J_0 [e^{qV_A/kT} - 1]$$

Since $V_A \rightarrow (-V_R)$ then $[e^{qV_R/kT} - 1] \rightarrow 0$

$$J_R = J_0$$

$$J_0 = AT^2 e^{-q(\phi D/kT)} \quad (1.33)$$

This describes the reverse saturation current for the Schottky barrier rectifier.

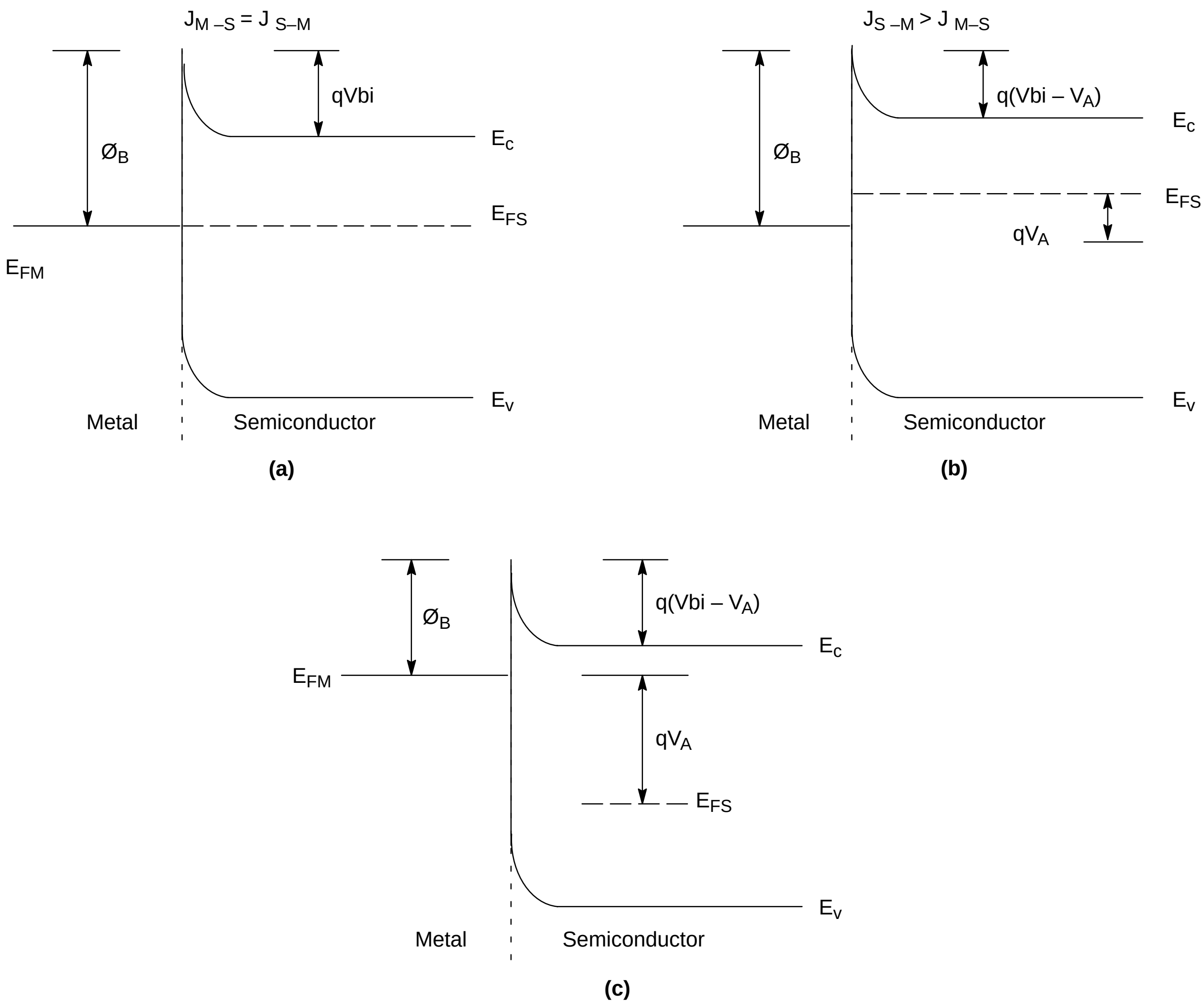


Figure 17. Schottky Barrier Rectifier Energy Band Diagrams (a) No Bias; (b) Forward Bias; (c) Reverse Bias

Similarly the diode equation for a pn junction obeys the general diode equation, which is:

$$J = J_0(e^{qV_A/kT} - 1) \tag{1.34}$$

where:

$$J_0 = q\left[\frac{D_P P_{no}}{Lp} + \frac{D_n n_{po}}{Ln}\right] \tag{1.35}$$

and where:

D_P = diffusion coefficient of holes

P_{no} = concentration of minority carrier holes in n–type material at equilibrium.

Lp = diffusion length of minority carrier holes in p–type material

D_n = diffusion coefficient of electrons

n_{po} = concentration of majority cater electrons in p–type material at equilibrium.

Ln = diffusion length of minority carrier electrons in n–type material.

Real Diode Characteristics

Forward Bias

Real diodes do not follow the ideal diode equation because of physical limitations of the device fabrication or design techniques. The bipolar silicon rectifier is fabricated using a p–n–n+ rectifier vertical structure, shown in Figure 18.

In this device the n–epitaxial layer is flooded with minority cater holes during forward bias. The bulk resistance is given by:

$$R = pl/A \qquad (1.36)$$

where:

p = resistivity
1 = drift region width
A = drift region area.

This bulk resistance is reduced during forward bias as a result. of the injection of minority carriers from the P^+ region. This allows the rectifier to carry high current density during forward conduction. This carrier action is called conductivity modulation. In a Schottky rectifier, the carrier action is with majority carriers and no conductivity modulation occurs. The charge transport mechanism in the Schottky is thermionic emission as opposed to diffusion and drift in the p–n–n+ structure of the bipolar rectifier. A cross–section of the vertical structure Schottky is shown in Figure 19 on the following page.

Typical forward characteristics of a 45 V Schottky and a 1000 V ultrafast rectifier are shown in Figure 20(a) and (b), respectively.

In the case of the Schottky rectifier, the forward voltage consists of four components:

$$V_f = \phi_B + (kT/Q)\, 1n\left(J_F/AT^2\right) + J_F Rsp_{epi} + J_F Rsp_{sub} \qquad (1.37)$$

where:

ϕ_B is the contribution due to barrier height in (eV).

(kT/q) ln (J_F/AT^2) is the contribution due to contact potential (V).

J_F RR_{epi} specific resistance of drift region voltage drop.

J_F RR_{sub} substrate resistance voltage drop.

J_F RR_{epi} and J_F RR_{sub} are the contributions due to the specific resistance of drift region and substrate, respectively.

In the case of a p–n rectifier, the forward voltage consists of four components also:

$$V_f = V_{bi} + J_F Rsp_{epi} + J_F Rsp_{sub} + J_F R_{lifetime} \qquad (1.38)$$

where:

V_{bi} the built–in potential described by equation 1.20

J_F Rsp_{epi} is specific resistance of drift region voltage drop

J_F Rsp_{sub} is specific resistance of substrate voltage drop

$J_F R_{lifetime}$ is specific resistance due to heavy metal lifetime reduction. In a standard recovery rectifier $J_F R_{lifetime}$, is zero.

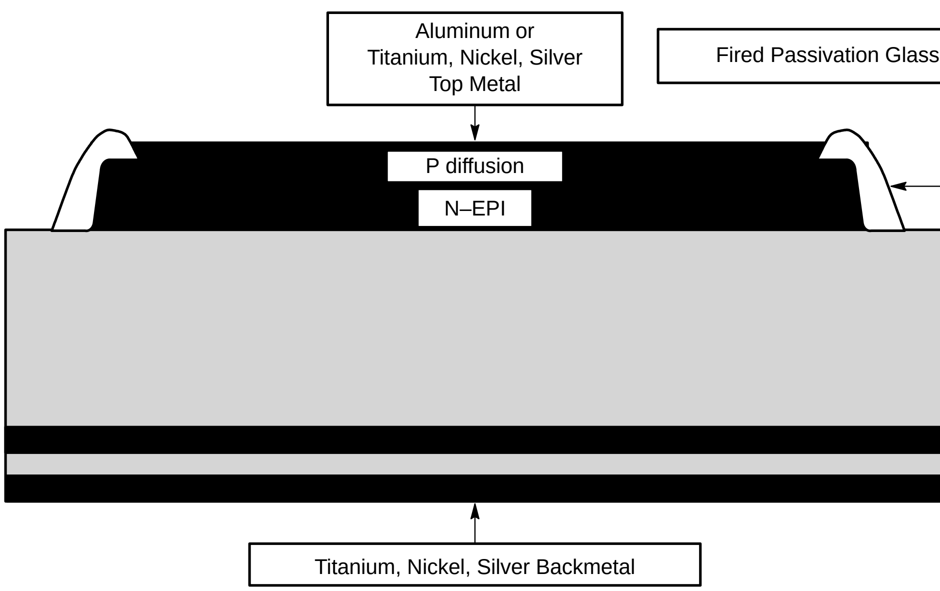


Figure 18. Vertical Structure of *pnn*+ Rectifier

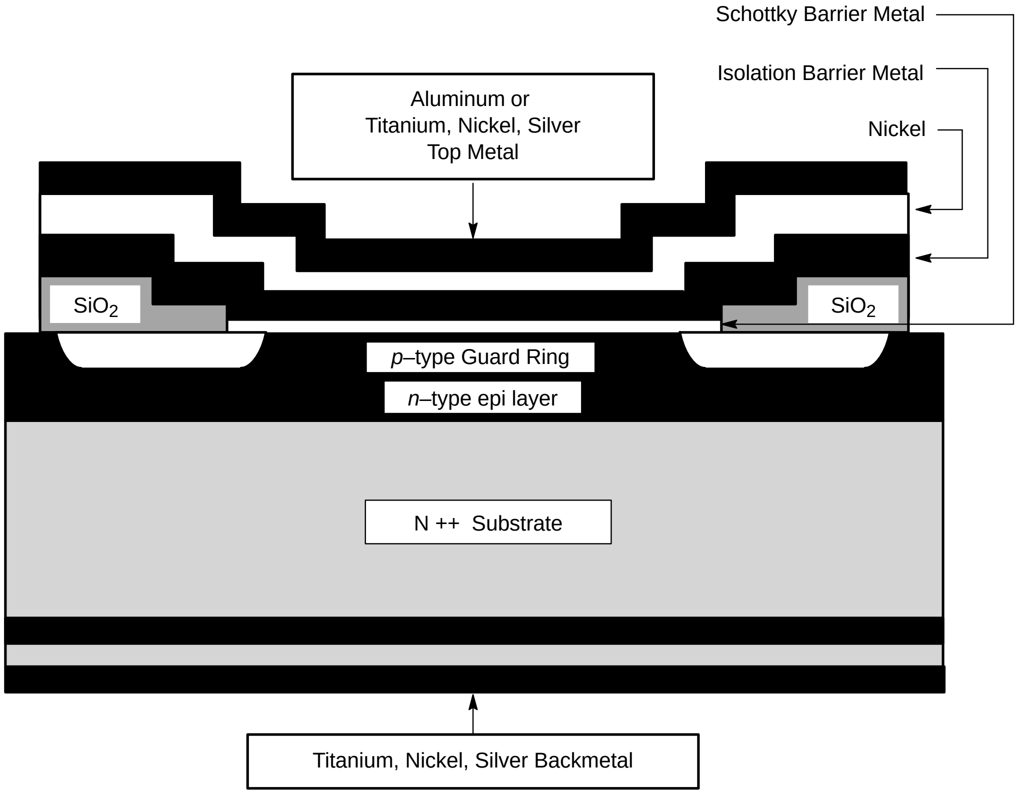


Figure 19. Vertical Structure of Schottky Rectifier

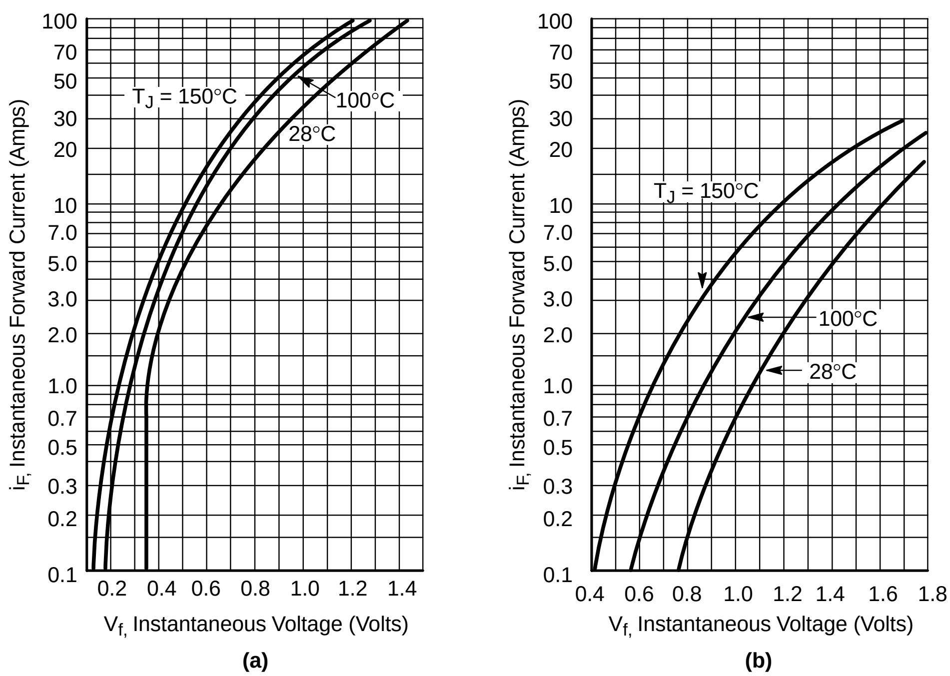


Figure 20. Typical Forward Voltage Characteristic
(a) Forward Voltage Characteristic of a 45 Volt Schottky MPR2O4SCT;
(b) Forward Voltage Characteristic of 1000 Volt Ultrafast MUR8100E

Reverse Bias

The reverse leakage current consist of four components for a bipolar p–n–n+ rectifier:

$$I_R = I_D + I_G + I_\tau + I_S \qquad (1.39)$$

where:

- I_D =diffusion current, sometimes called saturation current Jo
- I_G =space charge generation current
- I_τ =lifetime reduction current due to heavy metal doping
- I_S =surface passivation.

The hole diffusion current (I_D) in the drift region is minority carrier concentration in the drift region as shown in Equation 1.35. The space charge generation current (I_G) is the current that results from the charge transport mechanisms within the space charge region. The space charge region is the region that develops at the transition region between p–type and n–type materials. Under the conditions, of equilibrium described in Figure 21, hole–electron pairs continuously form in the space charge region as a result of thermal agitation.

However, the hole and electron forming a pair soon recombine. The period of time during which the electron–hole pair exists is referred to as the minority carrier lifetime. When a reverse bias is placed across the pn junction, the space charge region expands from the transition region into portions of the p and n materials on both sides of the junction. At the same time a larger field develops across the junction. The creation of this field, in turn, attracts mobile electrons and holes being generated in the space charge region and sweeps them into the p and n materials before they can recombine. This current is referred to as the space charge generation current I_G.

The lifetime reduction current (I_τ) is due to the influence of platinum or bold when introduced into the silicon crystal to enhance switching speed. The surface leakage current *(*I_S) is due to leakage paths on the surface of the semiconductor or charge influence in the insulation used to provide junction passivation In general, the contribution of surface passivation and lifetime reduction currents tends to be much higher than the space charge generation and diffusion components of the total reverse leakage current.

The reverse leakage current of a Schottky rectifier is dominated by the barrier height of the Schottky. In general Schottky reverse leakage currents are approximately five times higher than traditional Ultrafast rectifiers.

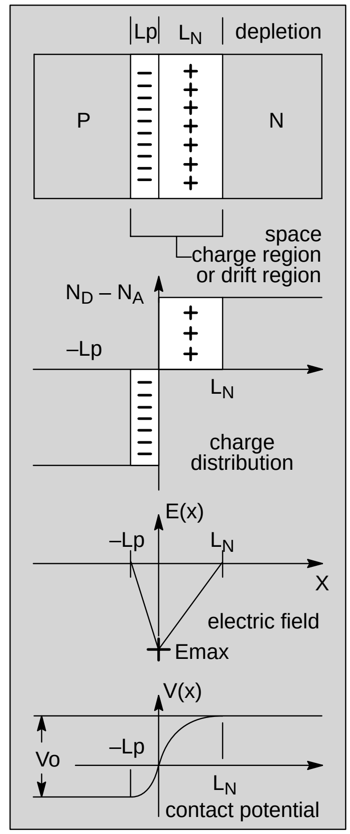


Figure 21. Property of *pn* Junction at Equilibrium Showing the Space Charge and Transition Region

Breakdown Voltage

Both Schottky and Ultrafast rectifiers exhibit avalanche breakdown. Avalanche breakdown is a process in which a single charge carrier ionizes an atom within the crystal lattice under the influence, of a strong electric field. Each electron and ion formed as a result of the ionization process is itself accelerated by the electric field through the space charge region. This acceleration imparts sufficient energy for the electron or ion to ionize other atoms it may impact and thereby create additional electrons and ions. The result of this rapid increase in reverse current is called avalanche breakdown.

For a step junction $p^+n^-n^+$ structure, Poisson's equation relates the electric field E to the charge density P in the depletion layer:

$$\frac{dE(\chi)}{d(\chi)} = \frac{P(\chi)}{\epsilon_S \epsilon_0} \qquad (1.40)$$

where:

ε_S = is the semiconductor dielectric constant and ε_0 is the permittivity of free space.

For a step junction where one side is heavily doped the electrical properties are determined by the characteristics of the lighter doped side. In $p^+n^-n^+$ step junction, the maximum field can be found by integrating from the depletion edge to χ_d the junction.

$$E_{max} = \int_{-\chi_d}^{0} \frac{qN_D}{\epsilon_S \epsilon_0} d(\chi) = \frac{qN_D \chi_d}{\epsilon_S \epsilon_0} \qquad (1.41)$$

$$E = -\frac{dv}{d\chi} = \frac{qN_D \chi}{\epsilon_S \epsilon_0} \qquad (1.42)$$

Integrating once the potential difference across this one–sided step junction gives:

$$V = \frac{-qN_D}{2\epsilon_S \epsilon_0} \chi_d \qquad (1.43)$$

$$\chi_d = \sqrt{\frac{2\epsilon_S \epsilon_0}{qN_D}} \; |V| \qquad (1.44)$$

$$\chi_d = \sqrt{\frac{2\epsilon_S \epsilon_0}{qN_D}} \; |V_B - V_A| \qquad (1.45)$$

Substitute 1.45 into 1.40:

$$E_{max} = \sqrt{\frac{2qN_D}{\epsilon_S \epsilon_0}} \; |V_B - V_A|$$

When the maximum field exceeds a critical field for silicon, then avalanche breakdown occurs.

Capacitance

Under reverse bias conditions both Schottky and $p^+n^-n^+$ junctions exhibit a voltage–dependent capacitance relationship. This relationship is expressed as capacitance per unit area;

$$C = \sqrt{\frac{\epsilon_S \epsilon_0 qN_D}{2|V_B + V_A|}} \qquad (1.46)$$

Capacitance is significant in Schottky rectifier applications at high frequency.

Capacitance is a function of bulk concentration, area of junction and barrier height As the barrier height is lowered, the capacitance of a Schottky increases. As the active area of the Schottky junction is increased, the capacitance. increases. As the concentration of the bulk region is. decreased; the capacitance is decreased.

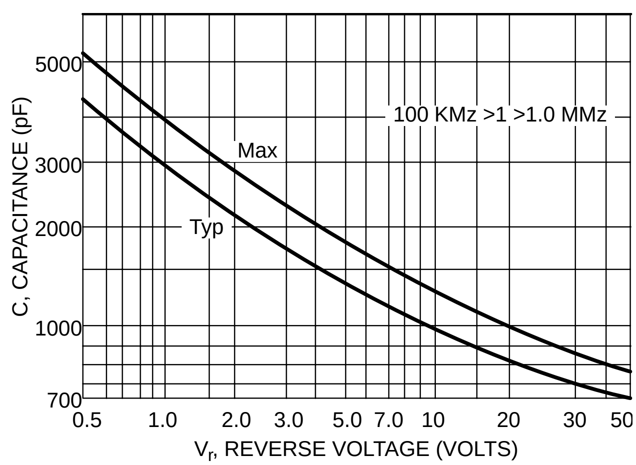


Figure 22. Capacitance Versus Reverse Voltage For Schottky Rectifier

Dynamic Switching

In general, Schottky technology offers switching performance because Schottky is a majority carrier technology. In fact, switching both forward and reverse directions are major drawbacks of $p^+n\ n^+$ technology. For example, when a $p^+n^+n^+$ rectifier is switched from reverse blocking to on–conduction as shown in Figure 23, its forward voltage drop exceeds its steady–state forward voltage drop. This phenomenon is called forward voltage overshoot during the turn–on transport. This is due to the fact that during Ugh–speed switching from the off–state to the on–state, current through the rectifier is limited by the maximum rate at which minority carriers are injected into the junction. A high voltage drop therefore develops across the diode for a short period of time until minority carriers diffuse into the junction and reduce the drift region resistance. This process was explained in an earlier section as conductivity modulation. The time it rakes for the diode to recover to within 10% of its steady–state forward conduction voltage is called the forward recovery time.

The reverse recovery of a $p^+n\ n^+$ rectifier is. an even more serious drawback than the forward recovery characteristic. Reverse recovery time is the time required for injected minority carriers to be removed from the space charge region when the device is “turning off.”’ The reverse recovery time. Which if we assume the reverse recovery characteristics are approximated to a triangle of base t_{rr} and height I_{rrm} peak reverse recovery current, then the reverse recovery charge Q_{rr} can be calculated to be:

$$Q_{rr} = 2I_{rrm} \cdot t_{rr} \text{ coulombs} \qquad (1.47)$$

Figure 24 shows the reverse, recovery characteristic of four power rectifier technologies.

The abrupt and soft recovery technologies of Figure 24 are traditional $p^+n^+n^+$ diodes with lifetime control for improved switching. However, when the switching frequency of power circuits increases, the turn–off di/dt also increases. This causes an increase in both the peak current I_{rrm}, and the resulting reverse recovery di/dt; this causes higher losses and more noise in the circuit.

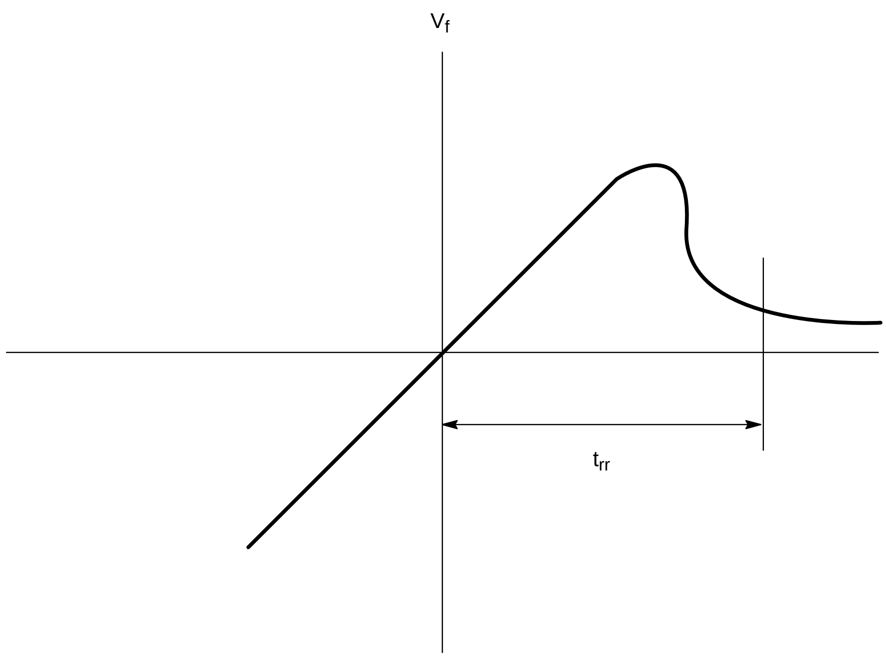


Figure 23. Forward Recovery Characteristics of *p+nn+* Diode When Switched From Reverse Blocking to On Conduction

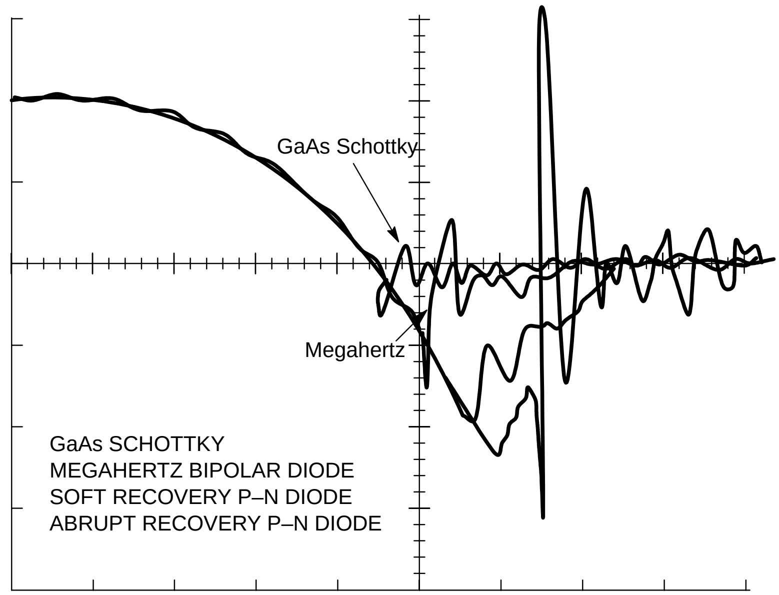


Figure 24. Reverse Recovery Characteristics

Rectifiers such as the Megahertz™ technology or the GaAs Schottky technology offer reduced t_{rr} and Q_{rr}. features and allow circuit operations at high frequency. Figure 25 shows the potential for Megahertz and GaAs Schottky rectifiers as enabling technologies for high–frequency applications.

Figure 26 shows the forward current relationship of two Schottky rectifiers at 25°C and 150°C. The phenomenon of negative temperature coefficient is observed.

From Figures 27 and 28 one can see that the gallium arsenide technology is essentially thermally stable since losses did not increase with temperature. whereas losses, although reduced, did increase with the Megahertz device over the 100°C, temperature excursion and a significant increase was exhibited by the Ultrafast rectifier.

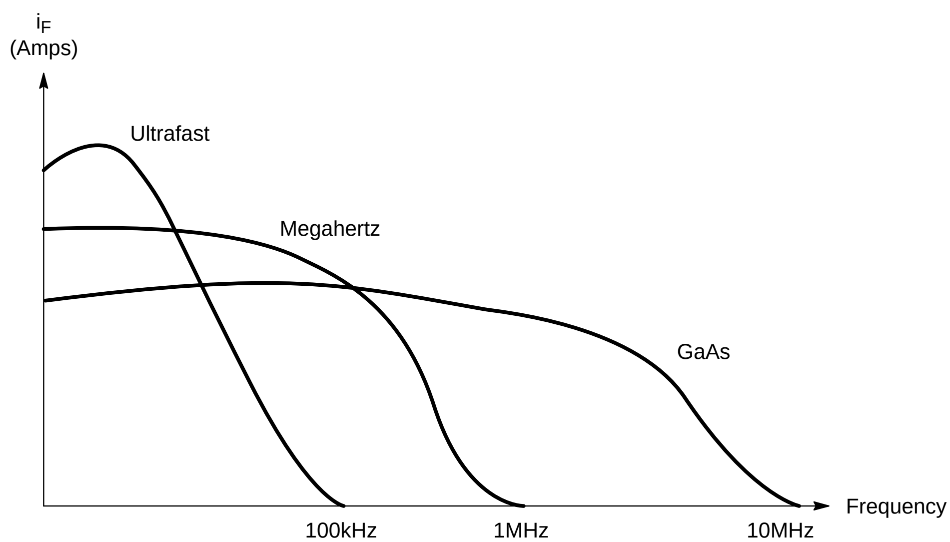


Figure 25. Safe Switching Area Curve for Ultrafast, Megahertz and Gallium Arsenide Schottky Power Rectifiers

Rectifier Applications

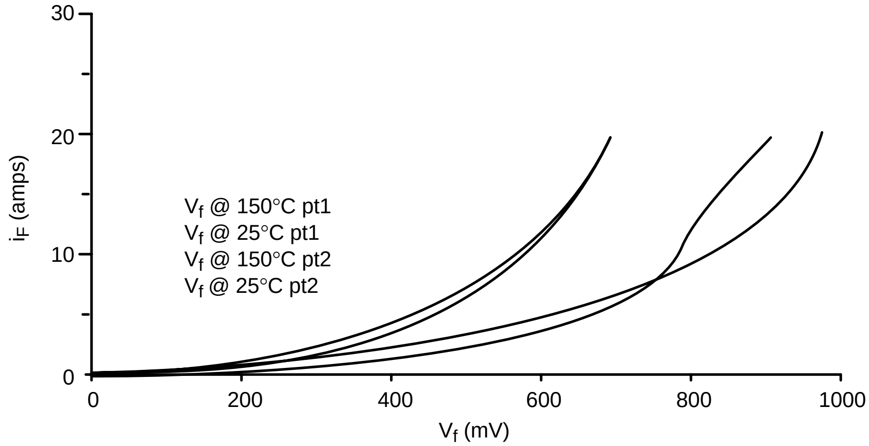


Figure 26. Forward Current Relationship of Two Schottky Rectifiers at 25°C and 150°C (the phenomenon of negative temperature coefficient is observed)

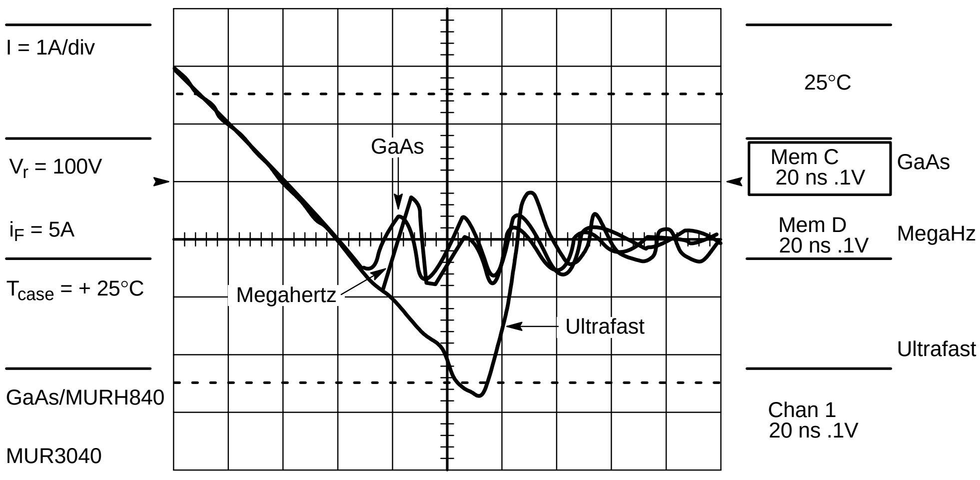


Figure 27. The Reverse Characteristic of Ultrafast, Megahertz and GaAs at T_A = 25°C. All Devices Exposed to the Same Test Conditions.

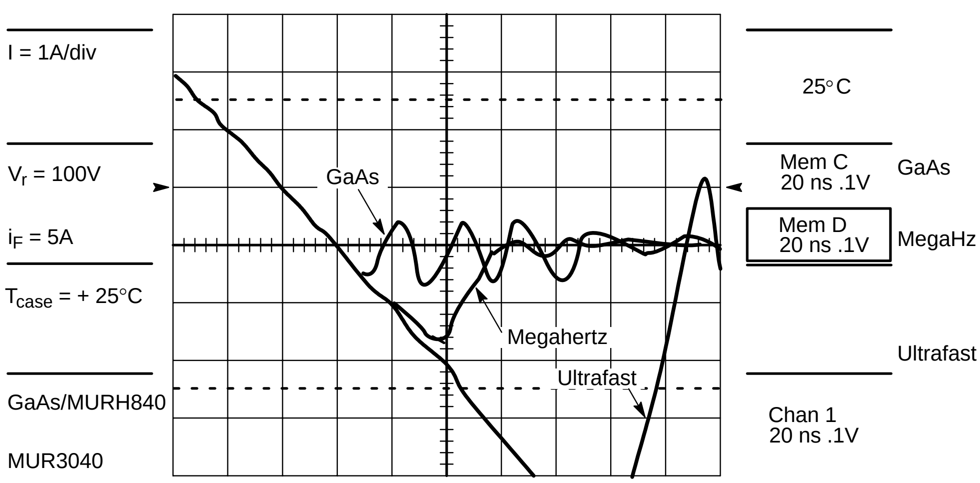


Figure 28. The Reverse Recovery Characteristic of Ultrafast, Megahertz and GaAs at T_A = 125°C. All Devices Exposed to the Same Test Conditions.

Silicon Temperature Coefficient Formula for Forward Voltage

When current is held constant, changes in forward voltage as a result of changes in junction temperature may be found from

$$V_f(T_j) = V_f(25^\circ) + \phi(T_j - 25^\circ) \qquad (1.48)$$

where

- $V_f(T_j)$ = forward voltage at T_j
- $V_f(25^\circ)$ = forward voltage 25°C
- ϕSi = temperature coefficient for silicon
- T_j = junction temperature

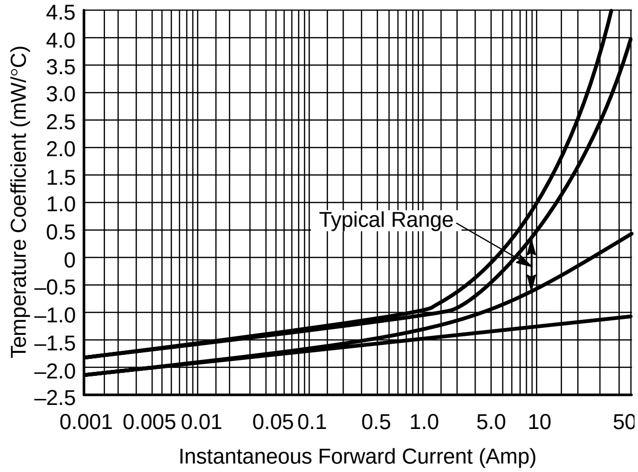


Figure 29. ∅$_{si}$ of Forward Voltage for a Family of Si Diodes

Design Tradeoffs

Table 3. Effects of Physical Change on Electrical Characteristics

Physical Change	Electrical Parameter Change										
	Increase						Decrease				
Increase Resistivity (Epi)		V_f	BV				t_{fr}		CT	I_r	t_{rr}
Area				CT	t_{rr}	I_r	t_{fr}	V_f			
Thickness(Epi)	t_{fr}	V_f	Bv	CT	t_{rr}						
Barrier Height		V_f							CT	I_r	
Lifetime					t_{rr}			V			

t_{lr} – Forward Recovery
CT – Capacitance Transient
t_{rr} – Reverse Recovery
V_f – Forward Voltage
BV – Breakdown Voltage
I_r – Reverse Leakage

Conclusion

Power rectifiers are essential components in the power electronics industry. As power rectifiers are improved in terms of conduction and switching performance they become enabling technologies for power conversion systems.

ON Semiconductor has established a technological strength, experience and skill resulting in functionally valuable products for our customers. This introductory chapter combines theory with description of real rectifier performance for both existing and emerging rectifier technologies.

At ON Semiconductor we have accepted the challenge from our customers to provide enabling power rectifiers and continue to strive towards the ideal device.

References

1. J.G. Irvin, *Resistivity of Bulk Silicon and of Diffused Layers in Silicon*, Bell System Tech J. Vol. 41, p. 387, 1962.
2. S.M. Sze and J.C. Irvin, *Resistivity, Mobility and Impurity Levels in GaAs, Ge and Si at 300°K*, Solid State Electronics, 11; 599 (1968).
3. Sze, *Physics of Semiconductors.* page 364–368.

Chapter 2

Basic Thermal Properties of Semicondutors

Basic Thermal Properties of Semiconductors

Three basic processes play a part in the removal of heat from the rectifier junction to the ambient air: (1) conduction (heat traveling through a material); (2) convection (heat transfer by physical motion of a fluid); and (3) radiation (heat transfer by electromagnetic wave propagation). Heat flows by conduction from the die to the package mounting surface in stud–, base–, or surface–mount pads, but it flows from the die through the leads to the mounting terminals in a lead–mounted part For case–mounted parts, convection and radiation are of primary importance in the design of the heat exchanger, which is covered in Chapter 13. For lead–mounted parts, radiation and convection from the body both play a role in removing heat from the die and are discussed later in this chapter. Transient thermal considerations and thermal runaway are often important design factors and receive treatment near the end of this chapter.

In order to simplify the analysis of heat flow, the concept of thermal resistance is used. Just as a material offers resistance to the flow of current, it may be thought of as offering resistance to the flow of heat. Resistance to heat flow is called thermal resistance and for steady–state conditions is given as:

$$R_\theta = \Delta T / P \qquad (2.1a)$$

or

$$\Delta T = R_\theta P \qquad (2.1b)$$

where:

R_θ = the thermal resistance in °C/W
ΔT = the temperature difference between points in °C
P = the power in watts.

Junction temperature of semiconductors must be held below the rating assigned to the part The junction is therefore commonly used for one of the reference points in applying Equation 2.1. The other reference point is the case for semiconductors enclosed in case–mounted packages or a specified point on a lead for semiconductors enclosed in a package intended for PCB insertion To denote the reference point, the symbols in Equation 2.1 have subscripts. Thus $R_{\theta JC}$ signifies thermal resistance, junction–to–case. While $R_{\theta JL}$ signifies thermal resistance, junction–to–lead. The corresponding temperatures are denoted ΔT_{JC} and T_{JL}, respectively.

Thermal Models

Thermal resistance may be used to form electrical models which permit calculation of the temperature rise at various points in a system. Similar to an electrical physical resistance, thermal resistance is not constant; changes in mounting, temperature, or power levels will cause some modification of values. Nevertheless, the concept provides a very valuable tool in handling thermal problems.

By use of a thermal model, complex thermal systems may be easily analyzed using electrical network theorems. The following sections discuss models for single chip case– and lead–mounted parts and for multiple chip assemblies.

Case–Mounted Rectifiers

The total thermal resistance, junction–to–ease, is composed of three identifiable thermal resistances, as shown in Figure 30. The die–bond thermal resistance is usually the largest value. Actual values are determined by the design of the device: the size of the chip, the type of the die bond, and the type and material of the package. Variations among parts from a given product line are the result of variations in the die–bond thermal resistance which is affected by the type of solder or bonding material used. As a general guide, however, thermal resistance as a function of the die area for various common diode packages using solder die bonds behaves as shown in Figure 31.

Some parts may have a piece of material inserted between the die and the package to take up stresses developed by differing thermal coefficients of expansion of the package and of the die. This technique allows a hard solder die attach technique to be used, which improves temperature cycling behavior Other parts may contain insulators that electrically isolate the chip from the package. These materials add another component of thermal resistance to the assembly.

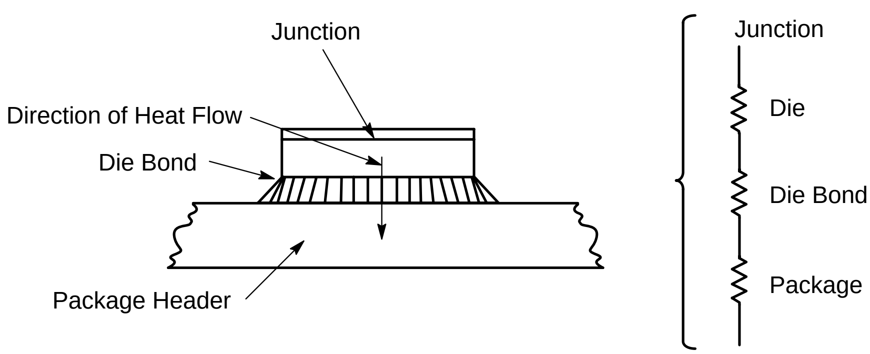


Figure 30. Thermal Resistance Components of the Junction–to–Case Thermal Resistance

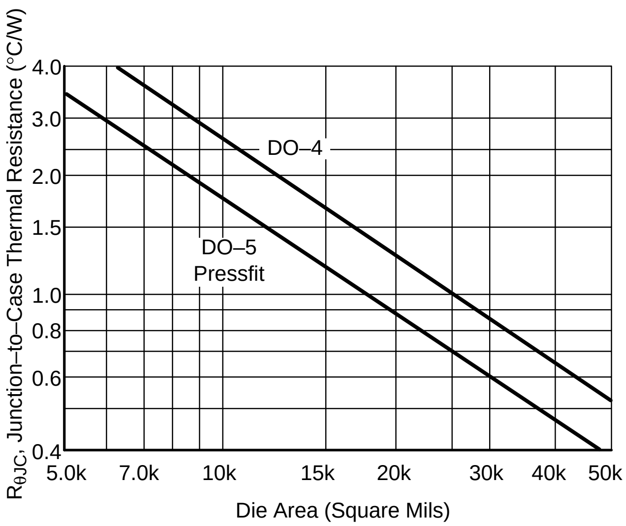


Figure 31. Typical thermal resistance of common rectifier packages having copper material in the heat flow path (data averaged from measurements at ON)

Thermal resistance (R_θ) follows the same general equation as does electrical resistance:

$$R_\theta = \rho \frac{l}{A} \qquad (2.2)$$

where:

- ρ = thermal resistivity
- l = length of thermal path
- *A* = area of thermal path

The equation states that thermal resistance is inversely proportional to area; however, the data of Figure 31 does not indicate this relationship exactly. The deviation is caused because the area of heat flow through the package is not the same as the die area. As heat flows, it spreads out toward the edges of the package; consequently, as larger die are placed in a given package, the area for spreading reduces proportionally.

Lead–Mounted Parts

In the axial lead–mounted rectifier, heat travels down both leads to some kind of a heat dissipator, which is usually nothing more than a printed–circuit wiring pattern. Heat is also removed from the package by convection and radiation, which make the thermal circuit model immensely more complicated for a lead–mounted part than for a case–mounted part. However, certain lead–mounted parts are easily handled because the thermal resistance of both leads is identical and quite low compared to the package radiation and convection components which maybe neglected. Examples of parts in which this simplified approach is satisfactory are the MR750 series. Thermal resistance as a function of lead length is shown in Figure 32. Note that the thermal resistance is linearly proportional to lead length, indicating that the package heat transfer components play a negligible role in the total thermal resistance. If the package thermal resistance components were not negligible, the lines would curve as the lead length increased.

Data is often given for the case where both leads have identical lengths. However, identical lead lengths will not result in lowest thermal resistance to the mounting points since the net thermal resistance is composed of two parallel paths. The lowest net value will always occur when one of the paths is made as short as possible. For example, suppose a mounting situation is encountered where the leads must take up a 1 inch span. If each lead were 1/2 inch long, the thermal resistance (from Figure 32) is l3°C/W maximum. However, the device could be mounted with one lead 1/8 inch long and the other 7/8 inch long. The thermal resistance from junction to the end of each lead is 4°C/W for the 1/8 inch lead and 23°C/W for the 7/8 inch lead. The net thermal resistance of the parallel combination is 3.4°C/W. The reduction from 13°C/W is quite significant but to take advantage of this reduction the mounting terminal must have a low thermal resistance to the ambient As the span becomes less, the advantages of asymmetrical mounting become less significant.

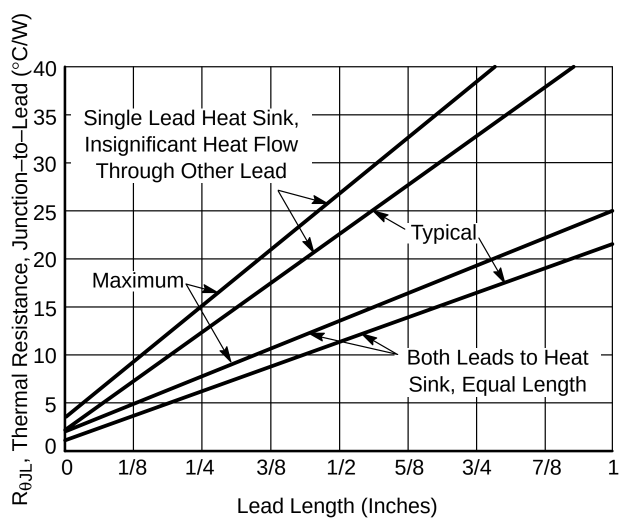
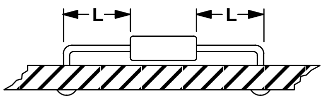


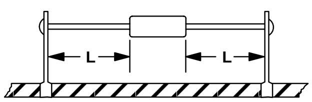
Figure 32. Thermal Resistance as a Function of Lead Length for MR751 Series Axial–Lead Rectifiers

As a design guide, when using lead–mounted pans, Figure 33 shows typical data for three popular case types. The data should not be taken as absolute because junction–to–ambient thermal resistance cannot be regarded as a design constant. The factors involved are discussed in depth in Chapter 13.

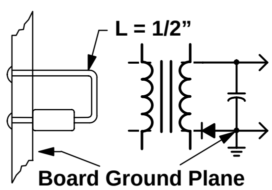
Parts with asymmetrical lead conduction and/or significant convection and radiation from the case requite use of a complete thermal model. Figure 34 shows a satisfactory approximation.



MOUNTING METHOD 1
P.C. Board Where Available Copper Surface area is small



MOUNTING METHOD 2
Vector Push–In Terminals T–28

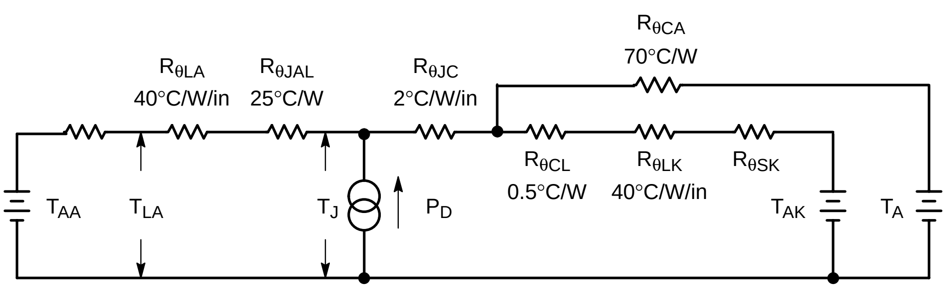


MOUNTING METHOD 3
P.C. Board with Copper Surface of Area A

CASE #	MOUNTING METHOD	LEAD LENGTH L			
		1/8″	1/4″	1/2″	3/4″
59 (DO–41)	1	65	72	82	92
	2	74	81	91	101
	3	40°C/W (L = 3/8″, A = 2.25″ ^{2})			
60	1	–	55	–	58
	2	–	65	–	68
	3	25°C/W (L = 5/8″, A = 6.25″^{2})			
267	1	50	51	53	55
	2	58	59	61	63
	3	28°C/W (L = 1/2″, A = 6.25″^{2})			

Data shown for the mountings shown is to be used as typical guideline values for preliminary engineering or in case the tie point temperature cannot be measured.

Figure 33. Typical Values for $R_{\theta JA}$ in Still Air



Use of the above model permits calculation of average junction temperature for any mounting situation. Lowest values of thermal resistance will occur when the cathode lead is brought as close as possible to a heat dissipator, as heat conduction through the anode lead is small. Terms in the model are defined as follows:

TEMPERATURES

T_A = Ambient
T_{AA} = Anode Heat Sink Ambient
T_{AK} = Cathode Heat Sink Ambient
T_{LA} = Anode Lead
T_{LK} = Cathode Lead
T_J = Junction

THERMAL RESISTANCES

$R_{\theta CA}$ = Case to Ambient
$R_{\theta SA}$ = Anode Lead Heat Sink to Ambient
$R_{\theta SK}$ = Cathode Lead Heat Sink to Ambient
$R_{\theta LA}$ = Anode Lead
$R_{\theta LK}$ = Cathode Lead
$R_{\theta CL}$ = Case to Cathode Lead
$R_{\theta JC}$ = Junction to Case*
$R_{\theta JAL}$ = Junction to Anode Lead (S bend)

*Case temperature reference is at cathode end.

Figure 34. Approximate Thermal Circuit Model for a Case 60 Part

Use of the thermal model of Figure 34 to calculate junction temperature is illustrated by the following example:

Cathode Lead Length = 1/4 inch T_A = 60°C

Anode Lead Length = 1/2 inch T_{AA} = 70°C

$R_{\theta SA} = R_{\theta SK}$ = 40°C/W (typical for printed board wiring)

T_{AK} = 80°C

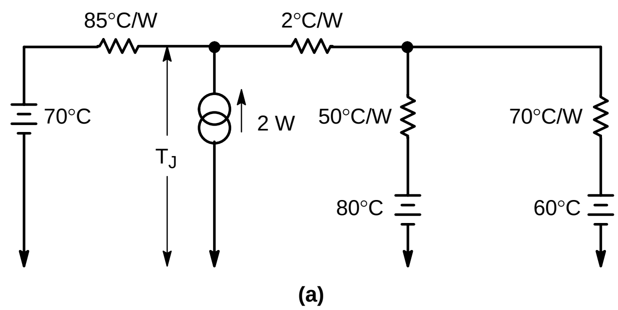
From the data in the figure, calculate:

$R_{\theta LA}$ = 40 x 1/2 = 20°C/W,

$R_{\theta LK}$ = 40 x 1/4 = 10°C/W.

The model of Figure 34 may be successively simplified by using Thevenin's network theorem as illustrated by Figure 35. The resulting junction temperature is 117°C. Thus, the effective thermal resistance, junction–to–ambient, is (117–60)/2 = 28.5°C/W; however, the number is not especially meaningful because the tempera–tufts of the ambient and the printed board wiring are not the same.

The concept of a single thermal resistance number to describe the thermal properties of axial–lead parts must be used with caution, because the ambient temperature is generally not the same as the temperature of the points which serve as a heat sink for the leads. Furthermore, in the case of an asymmetrically mounted part, there is little likelihood that either mounting point would be at the same temperature because more heat flows out through the low resistance lead. Thus, thermal resistance of a lead–mounted diode in terms of a single value is only useful when all the reference temperature points involved have the same temperature.



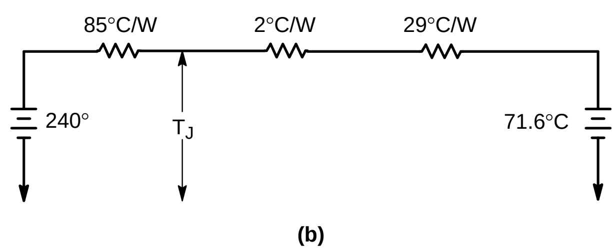


Figure 35. Successive steps in the solution of the example in the text. Thermal problems may be solved by applications of network theorems. (a) Model with example numbers inserted; (b) Simplified thermal circuit obtained by applying Thevenin's Theorem to each side of the model

Multiple Chip Devices

Assemblies having more than one die per package – for example full–wave, bridge, or three–phase configurations–have a considerably more complicated thermal situation because of thermal coupling between the die. To appreciate how the coupling problem can be handled, consider the thermal model of Figure 36. Coupling is represented by the resistances connecting all the die together. The coupling between adjacent die is primarily a result of heat transfer through the lead frame, but the surrounding encapsulate is also a thermal conductor. In addition, each die is also coupled to the diametrically opposite one through the molding in the center of the assembly. The center coupling requires the addition of two other resistance paths.

Since the model is so complex, it is easier to work in terms of a coupling factor. The general equation for thermally coupled die can be written as follows:

$$\Delta T_{J1} = R_{\theta 1}P_{D1} + R_{\theta 2}K_{\theta 2}P_{D2} + R_{\theta 3}K_{\theta 3}P_{D3} + \cdot\cdot\cdot R_{\theta n}K_{\theta n}P_{Dn} \quad (2.3)$$

where:

ΔT_{J1} = the rise in junction temperature of diode 1 with respect to the reference temperature

$R_{\theta n}$ = the thermal resistance of diodes 1 through n

P_{Dn} = the power dissipated in diodes 1 through n

$K_{\theta n}$ = the thermal coupling between diode 1 and diodes 2–4

n = the die of interest

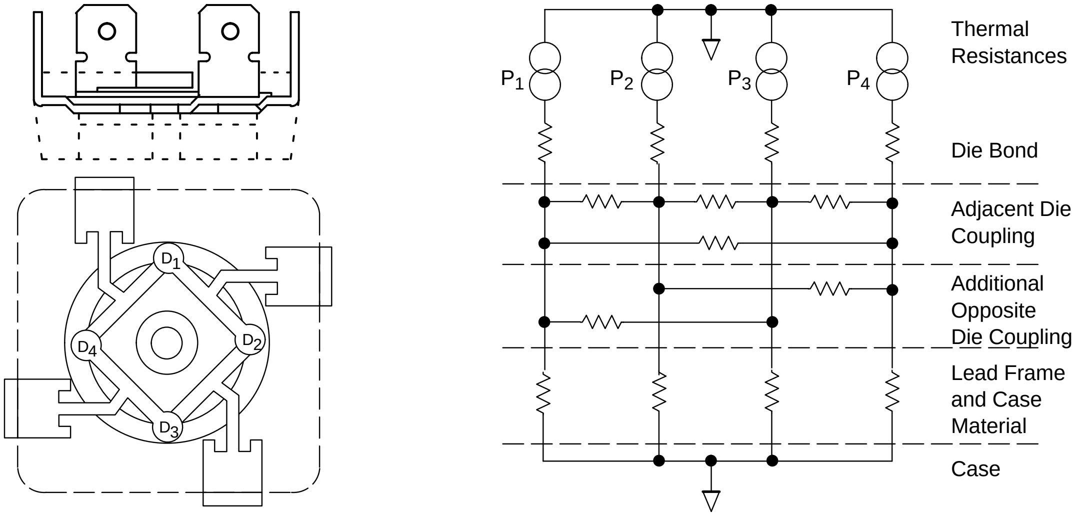


Figure 36. Sketch of Physical Construction and Thermal Model for a Bridge Assembly

Assuming equal thermal resistance for each die, Equation 2.3 simplifies to

$$\Delta T_{J1} = R_{\theta 1}(P_{D1} + K_{\theta 2}P_{D2} + K_{\theta 3}P_{D3} + \cdots K_{\theta n}P_{Dn}) \quad (2.4)$$

For the condition where $P_{D1} = P_{D2} = P_{D3} = P_{Dn}$, the total dissipation $P_{DT} = n\,P_{D1}$.

For this special case, Equation 2.4 can be further simplified to

$$\Delta T_{J1} = R_{\theta 1}(1 + K_{\theta 2} + \cdots K_{\theta n})\,P_{DT}/n \quad (2.5)$$

An effective package thermal resistance can be defined as follows:

$$R_{\theta(eff)} = \Delta T_{J1}/P_{DT} \quad (2.6)$$

Combining Equations 2.5 and 2.6 yields:

$$R_{\theta(eff)} = R_{\theta 1}(1 + K_{\theta 2} + K_{\theta 3} + \cdots K_{\theta n})/n \quad (2.7)$$

Assume an assembly has a thermal coupling coefficient between opposite dice of 0.12 and between adjacent dice of 0.20. If the individual die has a thermal resistance of 6.0°C/W, the effective bridge thermal resistance may be calculated from Equation 2.7 as

$$R_{\theta(eff)} = 6.0(1 + 0.20 + 0.120 + 0.20)/4 = 2.28^\circ C/W$$

This value is the one specified on the data sheet as the effective bridge thermal resistance, junction–to–case. Satisfactory average, steady–state temperature calculations may be made by multiplying $R_{\theta(eff)}$, by the total power dissipated in the package under a given load condition, provided the current in each diode is identical in wave–shape and amplitude.

The coupling factor is particularly valuable when a bridge assembly is operated in the split load circuit of Figure 37(b) instead of the more usual bridge circuit of Figure 37(a.) The following example illustrates the use of the data.

Assume that the previous assembly is used in the circuit of Figure 37(b) and operates such that

$$P_{D1} = P_{D2} = 10\ W$$
$$P_{D3} = P_{D4} = 5\ W$$

Maximum temperature rise occurs in diodes 1 and 2. Using Equation 2.4 and substituting values:

$$\Delta T_{J1} = 6.0\,[10 + (0.02)(10) + (0.12)(5) + (0.12)(5)] = 79.2$$

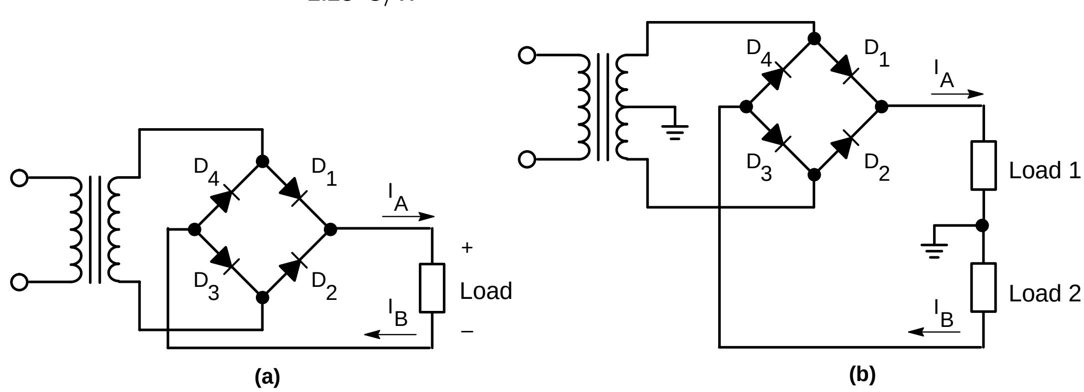


Figure 37. Basic Circuit Uses for Bridge Rectifiers (a) Standard Circuit; (b) Split Load Circuit

Transient Response of Semiconductors

Each component of thermal resistance in a semiconductor also has a thermal capacitance associated with it, related to the mass of the material. A high capacitance is desirable in order to improve the transient and overload capability of the part since it slows the response to a power pulse.

A mathematical analysis of one–dimensional heat flow indicates that the thermal response follows the relationship $T_J \propto \sqrt{t}$. However, the various time constants associated with a semiconductor affixed to a package, and the fact that heat flow is not usually one–dimensional, make the response extremely difficult to calculate. The most practical method of handling the transient thermal problem is to measure the thermal response of the semiconductor to a step of input power and to present the data in a graph. The thermal resistance as a function of time is called transient thermal impedance and may he calculated by:

$$Z_{\theta JR(t)} = r(t)\, R_{\theta JR} \qquad (2.8)$$

where:

$r(t)$ = fraction of steady state value at a given time

$R_{\theta JR}$ = thermal resistance, junction to a reference point

Choice of reference point depends on the type of part. For case–mounted parts, the logical reference is the case. For lead–mounted parts the leads are generally chosen; however, the ambient ins sometimes used.

Thermal Response

Thermal response of lead–mounted semiconductors is difficult to present because of the effect of lead length. Figure 38 shows the thermal response of two lead–mounted parts. In general, the slope, between 10 and 100 ms, as on Figure 38(b), is followed until steady–state conditions are approached. Figure 38(b) shows transient thermal impedance, $Z_{\theta JL(t)}$. Where the effects of lead length are considered. Curves for other lead lengths can be sketched using the given curves as a guide and flattening the curve at the appropriate value of steady–state thermal resistance.

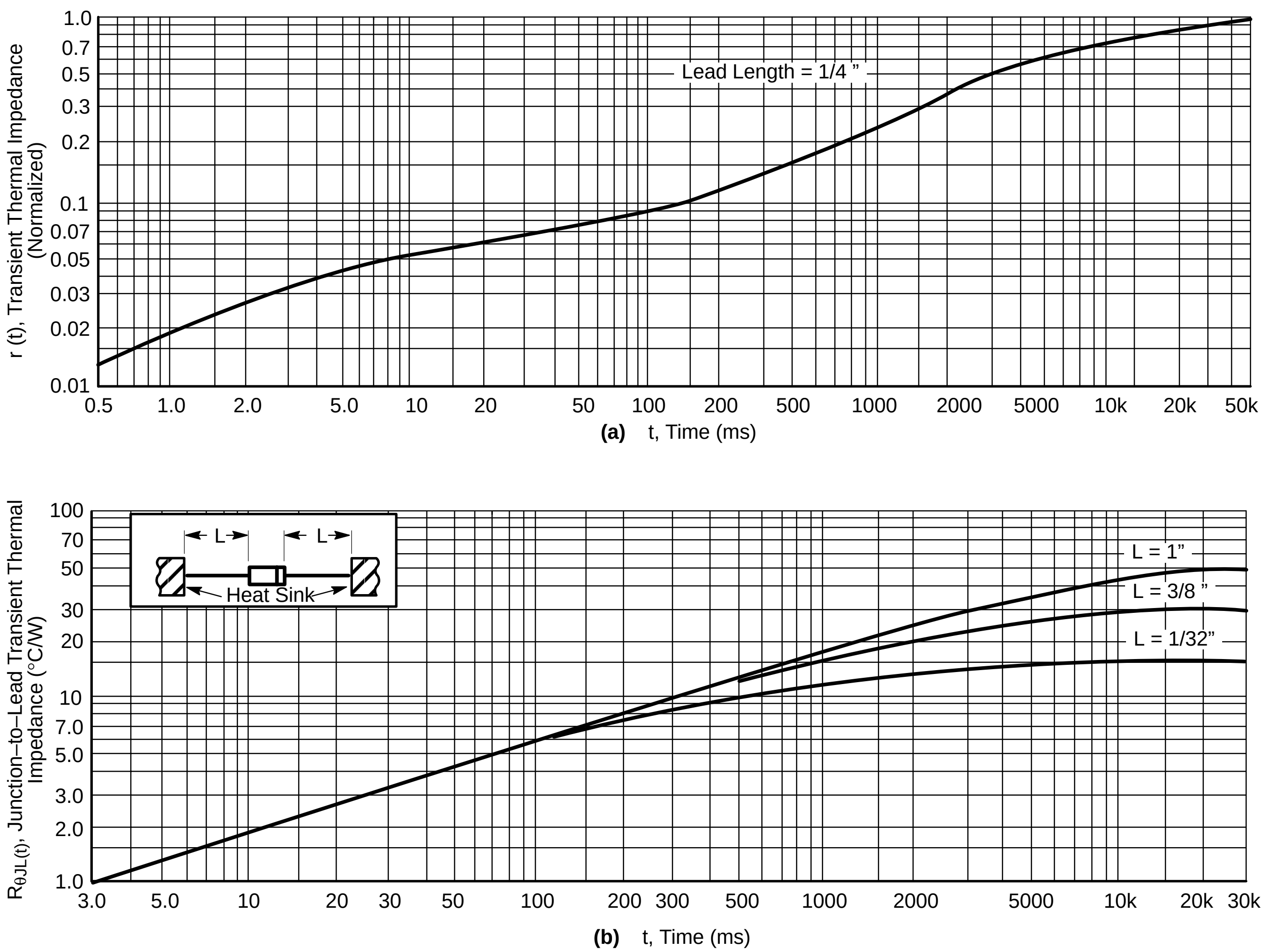


Figure 38. Typical Thermal Response of Lead–Mounted Rectifiers
(a) Case 59 with Fast Recovery Diode, Plotted Normalized;
(b) Case 59 with Standard Diode, Plotted in Terms of °C/W

Figure 39 shows response curves for typical power rectifiers in stud–mount and press fit cases. Differences in dies and manufacturing techniques make it difficult to quantitatively explain behavior, although it is evident that the more massive packages and larger die have slower responses.

Measurements of various rectifier diode arrays have revealed that times well ever 10 seconds are required for temperature changes of other die to become manifest at one of the other die in the same package. As a result, transient coupling can be neglected because of the relatively high frequencies employed in electronic work.

When rectifiers are used in intermittent operation the thermal response of the heat sink may be used to advantage. Heat sink properties are discussed in Chapter 13.

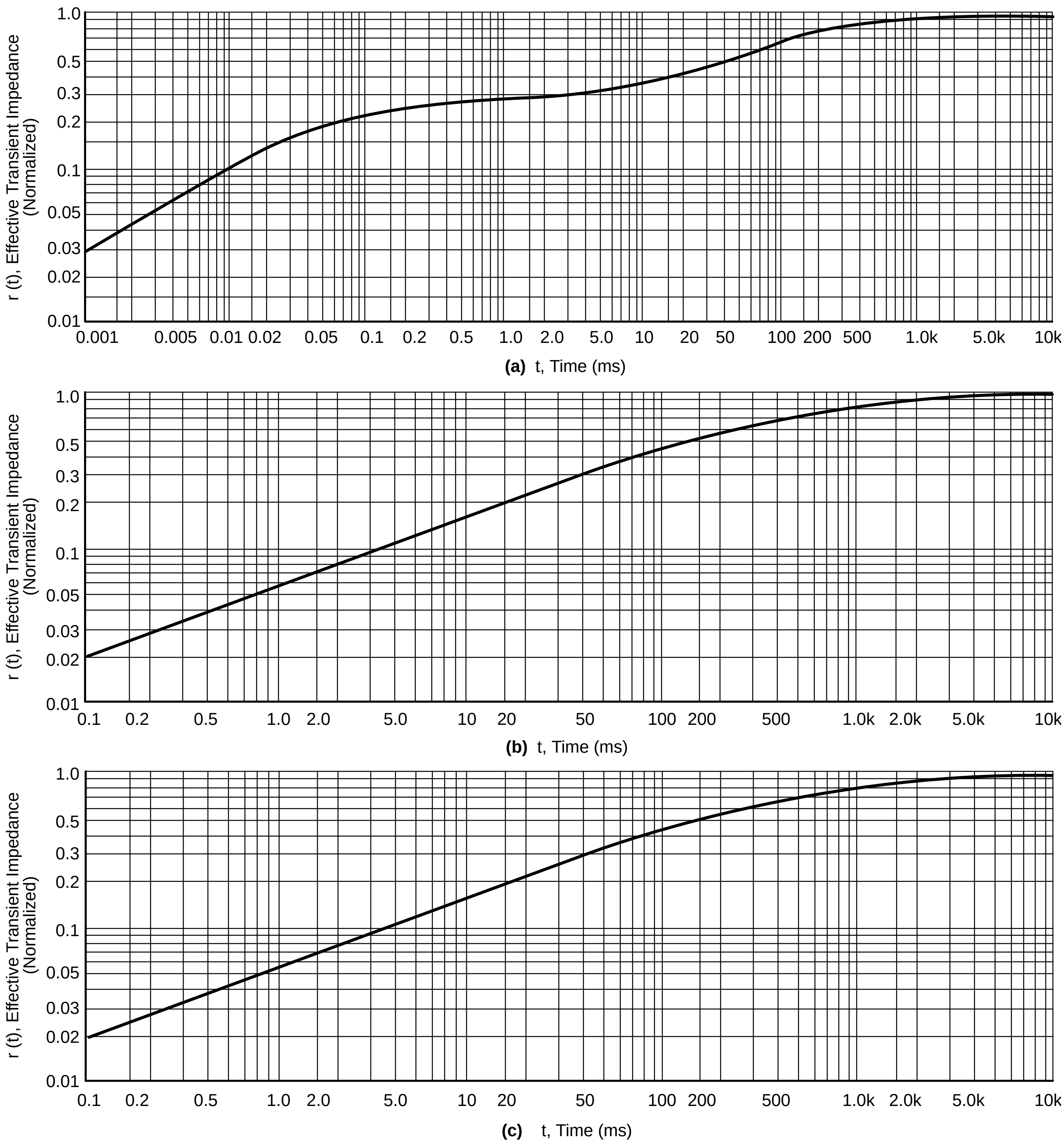


Figure 39. Typical Thermal Response of Stud–Mounted and Press–fit Rectifiers (a) DO–4 Package; (b) DO–5 Package; (c) DO–21 Package

Use of Thermal Response Data

A simple equation to permit calculation of temperature for arbitrary pulse trains with random variations is impossible to derive. However, since the heating and cooling response of a semiconductor is essentially the same, the superposition principle may be used to solve problems which otherwise defy solution. Using the principle of superposition, each power interval is considered positive in value and each cooling interval negative, lasting from time of application to infinity. By multiplying the thermal resistance at a particular time by the magnitude of the power pulse applied, the magnitude of the junction temperature change at a particular time is obtained. The net junction temperature is the algebraic sum of the terms.

The application of the superposition principle is easily seen by studying Figure 40. Part (a) illustrates the applied power pulses. Part (b) shows these pulses transformed into pulses lasting from time of application arid extending to infinity; at t_0, P_1 starts and extends to infinity; at t_1, a pulse ($–P_1$) is considered to be present and thereby cancels P_1 from time t_1, and so forth with the other pulses. The junction temperature changes, due to these imagined positive and negative pulses., are shown in part (c). The actual junction temperature is the algebraic sum as shown in part (d).

Problems may be solved by applying the superposition principle exactly as described; the technique is referred to as the pulse–by–pulse method. It yields satisfactory results when the total time of interest is much less than the time required to achieve steady–state conditions. This method must be used when an uncertainty exists in a random pulse train as to which pulse will cause the highest temperature.

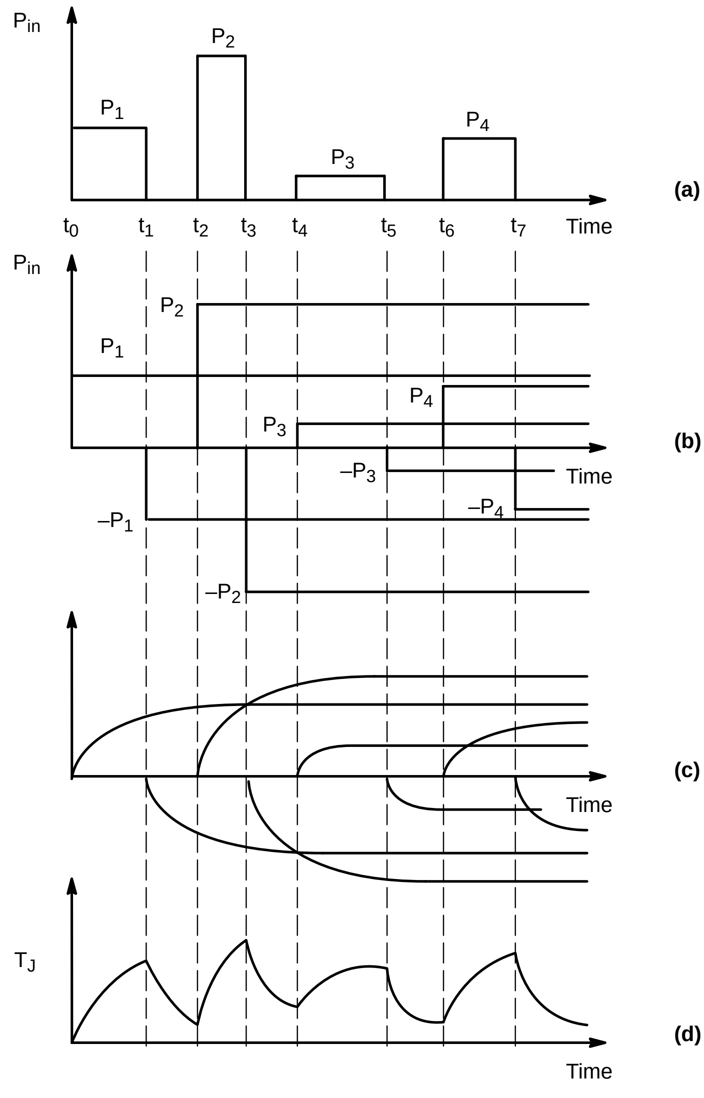


Figure 40. Application of Superposition Principle (a) Input Power; (b) Power Pulses Separated into Components; (c) T_J Change Caused by Components; (d) Composite T_J

Where power surges or overloads occur with uniform trains of repetitive pulses, more accurate answers with less work are obtained by using a model based on averaging the power pulses. The basis of this method is shown in Figure 41. The average power pulse is followed by one pulse of the repetitive train and then the overload pulse. The temperature is calculated at the end of the surge or overload.

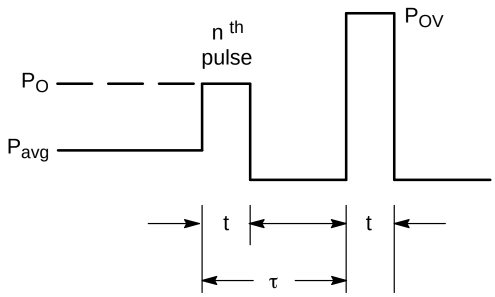


Figure 41. Model for a Repetitive Equal Pulse Train Having an Overload Condition

Handling Nonrectangular Pulses

Thermal response curves, such as those shown in Figure 38 and 39, are based on a step change of power; the response will not be the same for other waveforms. Thus far in this treatment a rectangular–shaped pulse has been assumed. It would be desirable to be able to obtain the response for any arbitrary waveform, but the mathematical solution is extremely unwieldy. The simplest approach is to make a suitable equivalent rectangular model of the actual power pulse and to use the given thermal response curves; the primary rule to observe is that the energy of the actual power pulse and the model are equal.

Experience with various modeling techniques has lead to the following guidelines:

1. For a pulse that is nearly rectangular, a waveform having an amplitude equal to the peak of the actual pulse, with the width adjusted so the energies are equal, is a conservative model (see Figure 42(a)).
2. Sine wave and triangular power pulses model well with the amplitude reduced by a factor. F_A, of 91% and 71%. respectively, of the peak and the width adjusted to 70% of the baseline width (as shown in Figure 42(b)). A power pulse having a $\sin^2$ shape models as a triangular waveform.

Rectifier diode forward power pulses are a combination of sin and $\sin^2$ waveforms if the current is a sinewave. In general, suitable mudding results if the amplitude waveform factor, F_A, is about 0.91 when the peak current is low and 0.71 when the peak current is high, that is, when the voltage drop across the forward resistance (R_F) is significant. (At low levels, diode voltage is fairly constant over the forward current cycle yielding a sine wave power waveform, while at high power levels the diode voltage is more nearly proportional to the instantaneous value of current resulting in a power waveform closer to a $\sin^2$ function. See Chapter 4.)

Rectifiers used with SCRs in phase control circuits must handle the current waveforms of part (c) of Figure 42. Empirically developed modeling rules for SCR power losses are indicated; the rules are slightly more conservative when used with rectifiers than with SCRs.

Power pulses having more complex waveforms are modeled by using two or more pulses as shown in Figure 42(d).

Note: A point to remember is that a high–amplitude pulse of a given amount of energy will produce a higher rise in junction temperature than will a lower amplitude pulse of longer duration having the same energy

The general equation for modeling is:

$$T_M = \frac{P_{(AV)}T_W}{F_A P_{(PK)}} \qquad (2.9)$$

where:
- T_M = pulse width of model
- P_{AV} = average power in the original waveform
- T_W = time duration of original waveform ($P_{(AV)}\,T_W$ = Area under the power waveform)
- F_A = amplitude waveform factor previously discussed
- $P_{(PK)}$ = peak power of original waveform

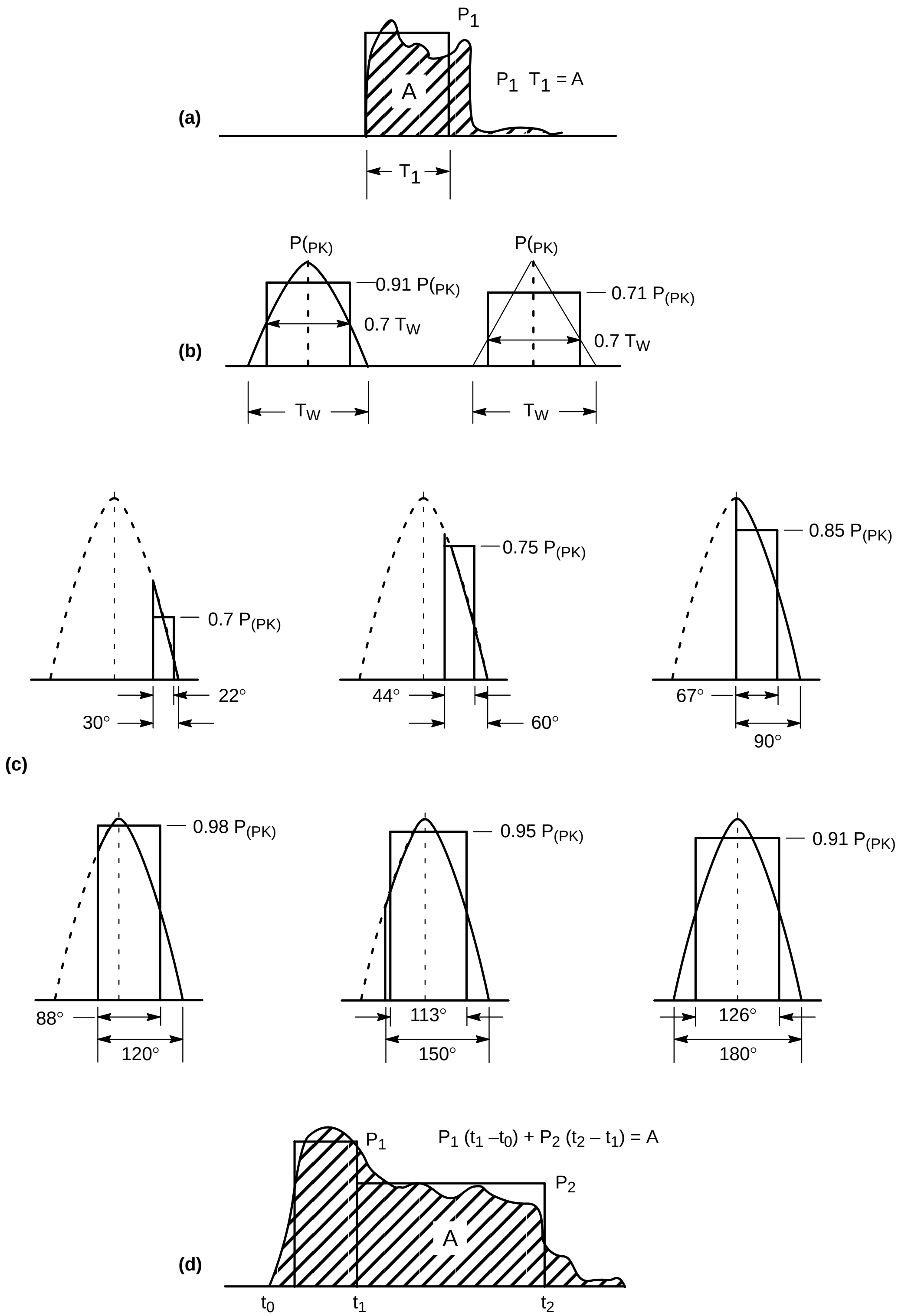


Figure 42. Models for Frequently Encountered Power Pulses
(a) Model for Nearly Rectangular Waveform; (b) Models for Sine Wave and Triangular Pulses;
(c) Models of SCR Current Waveforms for Various Conduction Angles (∝). Pulses for ∝ >120° are centered with respect to the sine wave. Pulses for ∝ > 90° start at the leading edge;
(d) Model for a Complex Waveform

Example Peak Temperature Calculations

The following examples illustrate proper analysis procedures for commonly encountered circuit applications. The first one is for a train of randomly spaced square pulses, and the second is for a single–phase circuit subjected to an overload. The second example requires use of proper power pulse modeling techniques to obtain a solution.

EXAMPLE 1

Finding T_J at the end of the nth pulse in a train of unequal amplitude, spacing, and duration:

General Equation:

$$T_n = \sum_{i=1}^{n} P_i[r(t_{2n-1} - t_{2i-2}) - r(t_{2n-1} - t_{2i-1})]R_{\theta JC} \quad (2.10)$$

where n is the number of pulses and P_i is the peak value of the ith pulse.

To find temperature at the end of each of the three pulses shown in Figure 43, Equation 2.10 becomes:

$$T_1 = P_1\, r(t_1)\, R_{\theta JC} \quad (2.10a)$$

$$T_2 = [P_1\, r(t_3) - P_1\, r(t_3 - t_1) + P_2 r(t_3 - t_2)]\, R_{\theta JC} \quad (2.10b)$$

$$T_3 = [P_1\, r(t_5) - P_1\, r(t_5 - t_1) + P_2 r(t_5 - t_2) - P_2 r(t_5 - t_3) + P_3 r(t_5 - t_4)]\, R_{\theta JC} \quad (2.10c)$$

Assume the following pulse conditions are applied to a rectifier with a transient thermal response as shown in Figure 38(a) and $R_{\theta JL}$ = 31°C/W for a lead length of 1/4 inch.

P_1 = 80 W	t_0 = 0	t_3 = 1.3 ms
P_2 = 40 W	t_1 = 0.1 ms	t_4 = 3.3 ms
P_3 = 70 W	t_2 = 0.3 ms	t_5 = 3.5 ms

Therefore:

$t_1 - t_0$ = 0.1 ms	$t_3 - t_1$ = 1.2 ms
$t_2 - t_1$ = 0.2 ms	$t_5 - t_1$ = 3.4 ms
$t_3 - t_2$ = 1 ms	$t_5 - t_2$ = 3.2 ms
$t_4 - t_3$ = 2 ms	$t_5 - t_3$ = 2.2 ms
$t_5 - t_4$ = 0.2 ms	

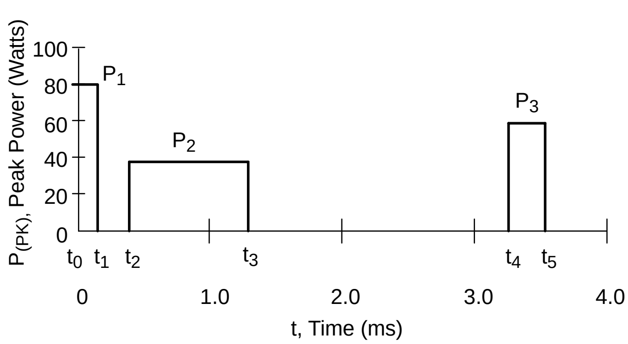


Figure 43. Pulse Train Analyzed for Peak Temperatures in Example 1 Text

PROCEDURE:

Find $r(t_n - t_k)$ for preceding time intervals from Figure 31a; then substitute into the set of Equations 2.10.

$$T_1 = P_1\, r(t_1)\, R_{\theta JL} = (80)(0.02)\,31 \approx 50^\circ C$$

$$\begin{aligned} T_2 &= [P_1\, r(t_3) - P_1\, r(t_3 - t_1) + P_2\, r(t_3 - t_2)]\, R_{\theta JL} \\ &= [80\,(0.058) - 80\,(0.056) + 40\,(0.054)]\,31 \\ &= [4.64 - 4.48 + 2.16]\,31 = 4.96 + 66.96 = 72^\circ C \end{aligned}$$

$$\begin{aligned} T_3 &= [P_1\, r(t_5) - P_1\, r(t_5 - t_1) + P_2\, r(t_5 - t_2) - P_2\, r(t_5 - t_3) + P_3\, r(t_5 - t_4)]\, R_{\theta JL} \\ &= [80\,(0.074) - 80\,(0.073) + 40\,(0.072) - 40\,(0.066) + 60\,(0.027)]\,31 \\ &= [(5.91 - 5.83) + (2.88 - 2.64) + (1.62)]\,31 \\ &= 2.48 + 7.44 + 50.22 = 60.14^\circ C \end{aligned}$$

The calculations above are performed so that, preceding the final answer, the residual temperature caused by the preceding pulses is evident. Note that very little residual temperature is left from the first pulse at the end of the second and third pulses. Also note that the second pulse gave the highest value of junction temperature, a fact not so obvious from inspection of Figure 43. Because of the high temperature caused by the second pulse, significant residual temperature from the second pulse was present at the end of the third pulse.

EXAMPLE 2

Find T_J at the end of an overload condition in a train of pulses of equal amplitude, spacing, and duration.

The overload–current condition shown in Figure 44 is applied to the rectifier used in the previous example and is modeled as shown in Figure 45. P_1 is the average power dissipation before the overload condition, P_2 is the average power dissipation during the overload condition, and P_3 is an equivalent peak power pulse resulting from the last overload pulse in the overload train. For the average current of 0.4 A, before the overload condition, and the average current of 2.2 A during the overload, the average power dissipation can be determined from rectifier power dissipation data. The data shows P_1 = 0.4 W and P_2 3.0 W for a resistive load. P_3 is a peak power and is obtained by noting that an average current of 2.2 A has a peak value of 6.9 A (3.14 x 2.2) causing a peak voltage of 1.45 V, obtained from rectifier V_F–I_F data. Using the modeling rules, the equivalent rectangular pulse, P_3, using a conservative waveform factor of 0.91 yields:

$$P_3 = (0.91)(6.9)(1.45) = 9.1\ W$$

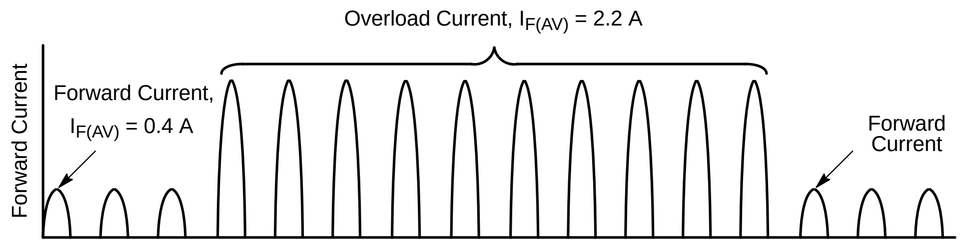


Figure 44. Overload Current Condition Used in Example 2 to Calculate Peak Junction Temperature

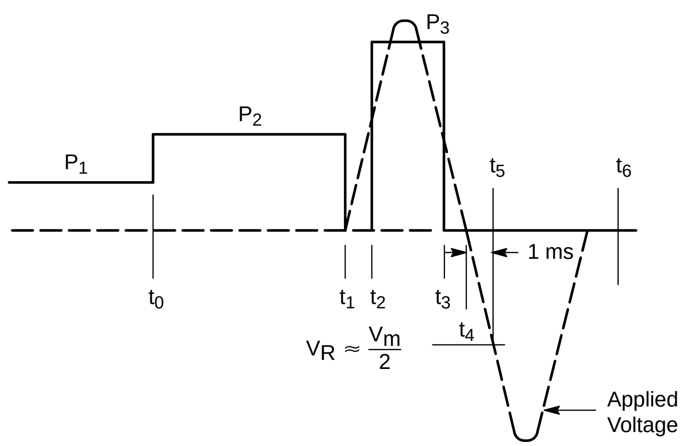


Figure 45. Model of Power Pulses Resulting from the Overload Current Conditions of Figure 44

The time $t_3 - t_2$ is found by adjusting the width T_M of the power pulse P_3 for equivalent energy by using Equation 2.9 to obtain

$$t_3 - t_2 = T_M = \frac{P_2 T_W}{P_3}$$

$$(t_3 - t_2) = \frac{(3)(16.667)}{9.1} = 5.5 \text{ ms}$$

thus,

$$(t_2 - t_1) = (t_4 - t_3) = \frac{8.34 - 5.5}{2} = \frac{2.84}{2} = 1.42 \text{ ms}$$

$t_1 = (9)(16.67) = 150$ ms $t_3 - t_1 = 6.92$

$t_2 = t_1 + (t_2 - t_1) = 151.42$ $t_5 - t_1 = 9.34$

$t_3 = t_2 + (t_3 - t_2) = 156.92$ $t_5 - t_2 = 7.92$

$t_5 = t_3 + (t_4 - t_3) + 1 = 159.34$ $t_5 - t_3 = 2.42$

The general equations used to calculate the junction temperature rise above ambient at t_3 and t_5 are as follows:

$$T_3 = [P_1 - P_1\, r(t_3) + P_2\, r(t_3) - P_2\, r(t_3 - t_1) + P_3\, r(t_3 - t_2)]\, R_{\theta JL} \quad (2.11a)$$

$$T_5 = [P_1 - P_1\, r(t_5) + P_2\, r(t_5 - P_2\, r(t_5 - t_1) + P_3\, r(t_5 - t_2) - P_3\, r(t_5 - t_3)]\, R_{\theta JL} \quad (2.11b)$$

Solving,

$$T_3 = [0.4 - 0.4\,(0.315) + 3\,(0.315) - 3(0.088) + 9.1\,(0.083)]\,31 = 1.71\,(31) \approx 53^\circ C$$

$$T_5 = [0.4 - 0.4\,(0.317) + 3(0.317) - 3(0.096) + 9.1\,(0.091) - 9.1\,(0.068)] = 1.15\,(31) \approx 36^\circ C$$

Thus, the junction temperature has cooled 17°C from time t_3 to time t_5. Reverse power dissipation, due to reverse voltage and leakage current, has been assumed negligible in the above example. If this is not the case, average reverse power dissipation is added to pulses P_1 and P_2*;* also a peak reverse power pulse is added between t_4 and t_5. Equations 2.11a and 2.11b are accordingly modified.

Thermal Runaway

Rectifier circuits may operate the rectifying diode such that thermal runaway is liable to occur and result in destruction of the diode. The problem arises because several characteristics are a function of temperature. Reverse recovery, reverse leakage, and usually forward power losses increase with temperature. As power is applied, junction temperature increases, causing power losses to increase. The additional power causes a further increase in temperature. Thus, a regenerative process is operating.

In any thermal system, the conditions for thermal stability are such that as the ambient temperature rises. the system must be capable of dissipating more heat than is generated. That is, the heat generated in the semiconductor must be less than the thermal conductivity from junction–to–air. Mathematically, the conditions for stability are

$$dP_D/dT_J < 1/R_{\theta JA} \qquad (2.12)$$

where dP_D/dT_J = change in power dissipation per unit change in temperature

$R_{\theta JA}$ = thermal resistance, junction–to–ambient.

A graphical approach [2] may be used to analyze the problem and illustrate the principles involved. It also permits visualizing the influence of ambient temperature and thermal resistance. Figure 46 shows how this is done; values are typical of those encountered with Schottky barrier power rectifiers, which have a sufficiently high reverse leakage current to cause thermal runaway.

Curve A is a plot of the heat generated at the rectifier junction versus junction temperature: The ordinate is simply $P_{R(AV)}$ and may be found from I_R vs. T_J and V_R curves; it is easy to do for dc conditions. The slope of curve A is dP_D/dT_J.

Curve B represents the power–dissipation capability (rate of heat flow) of the equivalent thermal circuit as a function of the junction temperature of the rectifier. The slope can be seen to be $1/R_{\theta JA}$, which corresponds to the thermal conductivity of the total thermal circuit from junction to ambient air.

With no voltage applied, the unit dissipates no power and the junction temperature is the same as the ambient. In Figure 46, this condition corresponds to the point T_A. When voltages are applied to the circuit, the heat generated by the semiconductor (in the form of dissipation) is P_1. This amount of internal heat generation in the semiconductor requires that the junction temperature increase to T_1. However, since the junction temperature has now been increased, the total power generated by the semiconductor must increase to P_2. In this manner, the junction temperature and resultant power generation increase until a stable condition is reached at point 1, where the power–generation and thermal–conductance curves intersect. The temperature at point 1 is labeled T_2 and may be read from the abscissa as approximately 48°C; the ambient temperature is at the intersection of curve B with the abscissa, and is 30°C for this example.

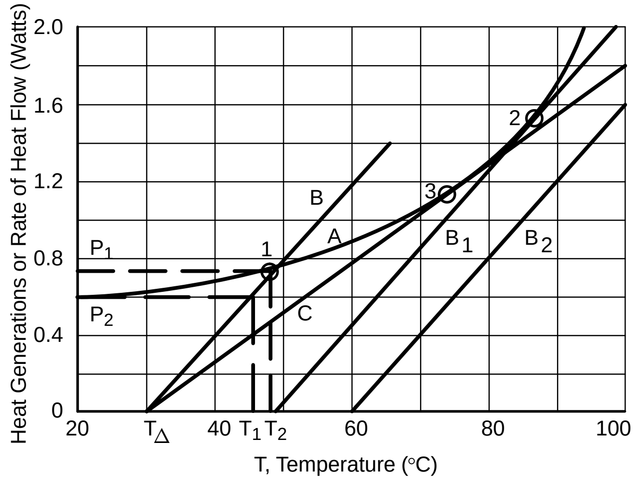


Figure 46. Graphical Analysis of Thermal Runaway

It is now possible to predict the effect of changing the ambient temperature. If the ambient temperature were increased. curve B would be translated to the right, resulting in a higher junction temperature because the intersection of curve A would be to the right of point 1. If the ambient temperature increased to 49°C (curve B_1), the junction temperature would be about 85°C. At values of ambient temperature below 49°C, for this particular thermal resistance (25°C/W), the circuit always fulfills the thermal stability criterion, $dP_D/dT_J < R_{\theta JA}$ and is therefore a thermally stable circuit. At an ambient temperature 49°C, curve B_1 is tangent to curve A at point 2, where $dP_D/dT_J = 1/R_{\theta JA}$. Under these conditions the circuit is conditionally stable, since a small incremental increase in junction temperature will cause an unstable condition. A further rise in the ambient temperature will cause the thermal conductance curve B_1 to fall below the power generation curve as shown by B_2. Under these conditions the semiconductor junction temperature cannot stabilize. The power generated within the semiconductor continues to increase in search of a stable condition until the junction is destroyed.

From this representation, the effect of changing the thermal conductance can be determined. If the cooling facility is changed (for instance, a better heat sink is used), the junction will run cooler for a given ambient temperature, since the slope will be increased. Curve B illustrates the result. Curve C shows what happens if $R_{\theta JA}$ is increased to 40°C/W. Note that the maximum ambient temperature allowed is 29°C and that the junction temperature is at 75°C at the point of thermal runaway (point 3).

Rectifier Applications

To aid the circuit designer in applying rectifiers having reverse leakage high enough to require that thermal runaway be considered in the design, a simplified approach has been devised and is presented on applicable ON Semiconductor data sheets. It is based on the relationship

$$T_{A(max)} = T_{J(max)} - R_{\theta JA} P_{F(AV)} - R_{\theta JA} P_{R(AV)} \qquad (2.13)$$

where $T_{A(max)}$ = maximum allowable ambient temperature
$T_{J(max)}$ = maximum allowable junction temperature (rated limit or the temperature at which thermal runaway occurs, whichever is lowest)
$P_{F(AV)}$ = average forward power dissipation
$P_{R(AV)}$ = average reverse power dissipation
$R_{\theta JA}$ = junction–to–ambient thermal resistance.

A reference temperature is defined by Equation 2.14:

$$T_R = T_{J(max)} - R_{\theta JA} P_{R(AV)} \qquad (2.14)$$

The reference temperature T_R may be limited only by $P_{R(AV)}$ or may be limited by thermal runaway. T_R is obtained by a computer solution of the basic stability criteria boundary given in Equation 2.12 as explained in the preceding discussion. T_R data is presented on graphs similar to that of Figure 47. With T_R known. $T_{A(max)}$ is found from

$$T_{A(max)} = T_R - R_{\theta JA} P_{F(AV)} \qquad (2.15)$$

Equation 2.15 is found by substituting Equation 2.14 into Equation 2.13.

Figure 47 shows how T_R is limited by both $T_{J(max)}$and thermal runaway. The transition between these two limitations is evident on the curves of Figure 47 as a difference in the rate of change of the slope in the vicinity of 115°C. The data is based upon dc conditions. For use in common rectifier circuits Table 4 indicates suggested voltage factors for an equivalent dc voltage to use for conservative design, i.e.

$$V_{R(equiv)} = V_{in(PK)} \times F_V \qquad (2.16)$$

The factor F_V is derived by considering the properties of the various rectifier circuits and the reverse characteristics of the particular diode. At a fixed junction temperature. the reverse power waveform is determined for various peak reverse voltage levels and average power is calculated. From $I_R - V_R$ data, the value of V_R is found which produces the same power dissipation as in the circuit; it is $V_{R(equiv)}$.

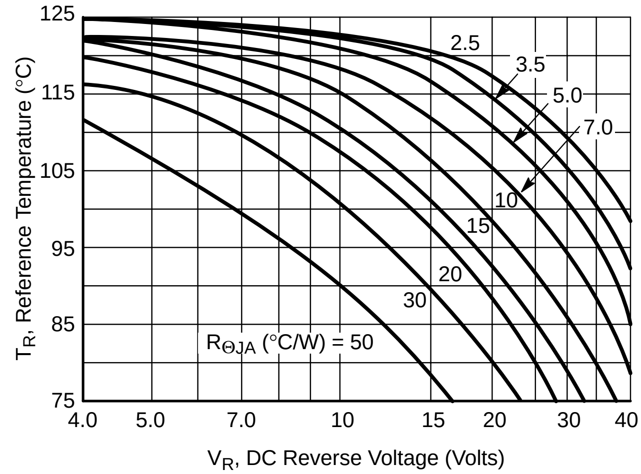


Figure 47. Maximum Reference Temperature for a Schottky Rectifier

EXAMPLE

Find $T_{A(max)}$ for the diode of Figure 47 operated in a 24 V DC supply using a bridge circuit with capacitive filter such that the load current $I_{L(DC)}$ = 10 A ($I_{F(AV)}$ = 5 A), $I_{F(PK)}/I_{F(AV)}$ = 0, input voltage = 20 V(rms,) $R_{\theta JA}$ = 7°C/W.

Step 1. Find $V_{R(equiv)}$. Read F_V = 0.65 from Table 4.

$V_{R(equiv)}$ = (1.41)(20)(0.65) = 18.3

Step 2. Find T_R from Figure 47. Read T_R = 108°C @ V_R = 18.3 and $R_{\theta JA}$ = 7°C/W

Step 3. Find $P_{F(AV)}$ from rectifier data. $P_{F(AV)}$ = 10 W at $I_{PK}/I_{(AV)}$ = 20 and $I_{F(AV)}$ = 5 A

Step 4. Find $T_{A(max)}$ from Equation 2.15.
$T_{A(max)}$ = 108–(7)(10) = 38°C.

References

1. Bill Roehr and Bryce Shiner, *Transient Thermal Resistance General Data and Its Use,* ON Semiconductor Application Note AN–569. ON Semiconductor Products Inc., Phoenix, Arizona.
2. Lloyd P. Hunter, *et al.*, *Handbook of Semiconductor Electronics,* Second Edition, Chapter 11. pp 11–80, 81. McGraw–Hill Book Co., Inc., New York, New York. 1962.

Table 4. Values for Factor F_V

Circuit	Half Wave		Full Wave, Bridge		Full Wave, Center Tapped*†	
Load	Resistive	Capacitive*	Resistive	Capacitive	Resistive	Capacitive
Sine Wave Square Wave	0.5 0.75	1.5 1.5	0.5 0.75	0.65 0.75	1.0 1.5	1.3 1.5

*Note that Vrwm = 2 $V_{in(PK)}$ *†Use line to center tap voltage for V_{in}.

Chapter 3

The SPICE Diode Model

The SPICE Diode Model

Special Acknowledgment
This chapter was written by Howard T. Russell, Jr., PhD, OPAL Engineering, Inc., August 4, 1991

Introduction

This chapter presents a description of the SPICE diode model and methods for the extraction of its parameters. A comprehensive examination of this model will be given along with comparisons of the characteristics of a real diode and those produced by the model. These comparisons will be used to illustrate the model's accuracy and limitations. Based on the nature of the model equations, mathematical methods for the extraction of the SPICE parameters (with the exception of the noise parameters) will be presented. These methods are easily implemented on a hand–held programmable calculator and will be used to extract the parameters for a ON Semiconductor MURHS4OCT rectifier as an example. Throughout this chapter, references are made to the characteristics and behavior of an ideal diode model. The development of this model may be found in several of the many well–known texts on the subjects of semiconductor device theory and pn junction diodes [1–4].

The SPICE model of a pn junction diode consists of mathematical equations, parameters, and variables, all of which are designed to work together to simulate as accurately as possible the electrical characteristics of a real device. The equations and variables used for the model in the SPICE program are fixed and not easily modified [5, 6]. Therefore, the accuracy resulting from a SPICE simulation of a diode depends on the precise extraction of its model parameters. Fortunately, there are several parameter extraction methods which can be applied to this model, some of which yield more accurate results than others. The inherent success of a particular method depends for the most part on how well device physics and theory are utilized in the design of the model, and upon the availability of data taken from a real device.

The SPICE Diode Model

The parameters for the SPICE model of the diode are given in Table 5 [5, 6]. The equations that use these parameters are divided into four model groups which are responsible for simulating various diode characteristics. These groups consists of the large–signal dc model, the small–signal ac model, temperature and area effects, and the noise model.

The Large–Signal DC Model

The large–signal behavior of the SPICE diode is characterized by the relationship between the dc current and voltage at its terminals. The parameters used to model this behavior are IS, RS, N, By, and 1EV. The parameter IS is the same as the reverse saturation current I_S for an ideal diode. The ohmic resistance RS is used to model the resistance of the metal contacts and the neutral regions under high–level injection. The emission coefficient N is used to modify the slope of the current versus voltage (I–V) characteristics curve. Finally, the parameters BV and IBV model the reverse breakdown behavior.

Figure 48 shows the equivalent circuit of the SPICE diode for large–signal dc analysis. This circuit contains an internal diode D_1, a series resistance having a value of RS, and a shunt conductance GMIN. The SPICE program adds this conductance, which is transparent to the user, around every internal pn junction to aid convergence. The program default value for GMIN is 10^{-12} mhos but can be set to any other non–zero value with a OPTIONS line in the circuit file [5].

Table 5. SPICE Diode Model Parameters

Name	Parameter	Default value	Typical value	Units
IS	Saturation Current	10.0 f	50.0 f	A
RS	Ohmic Resistance	0	2.0	ohm
N	Emission Coefficient	1.0	1.1	
TT	Forward Transit Time	0	10.0 n	sec
CJO	Zero–bias Junction Capacitance	0	10.0 p	F
VJ	Contact Potential	1.0	0.8	V
M	Junction Capacitance Grading Exponent	0.5	0.3	
EG	Energy Gap	1.11	1.11	eV
XTI	IS Temperature Exponent	3.0	3.0	
KF	Flicker Noise Coefficient	0	0.1f	
AF	Flicker Noise Exponent	1.0	1.0	
FC	CJ Forward–bias Coefficient	0.5	0.5	
BV	Reverse Breakdown	∞	100.0	V
IBV	Current at BV	1.0 m	200.0 p	A

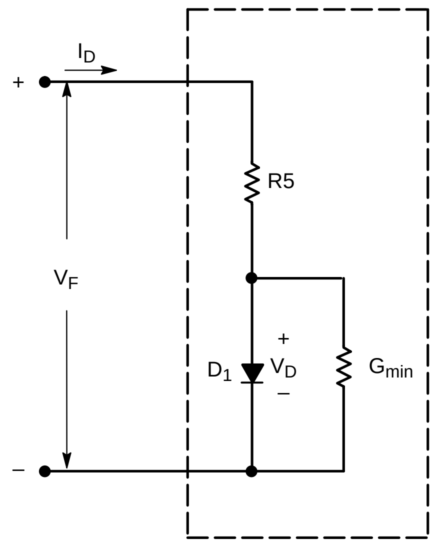


Figure 48. DC Large–signal SPICE Diode Model

The dc model variables consist of the voltage across the external diode terminals V_F. The voltage across the internal diode terminals V_D and the terminal current I_D. With these parameters and variables, the large–signal dc characteristics are modeled by the following equations

$$V_F = RS \cdot I_D + V_D \quad (3.1)$$

and

$$I_D = f(V_D) \quad (3.2)$$

where four regions of operation describe the functional relationship between the internal diode voltage and diode current.

(a) For $V_D \geq -5 \cdot N \cdot V_t$

$$I_D = IS \cdot \left\{ \exp\left(\frac{V_D}{N \cdot V_t}\right) - 1 \right\} + G_{MIN} \cdot V_D \quad (3.3)$$

(b) For $-BV < V_D - 5 \cdot N \cdot V_t$

$$I_D = -IS + G_{MIN} \cdot V_D \quad (3.4)$$

(c) For $V_D = -BV$,

$$I_D = -IBV \quad (3.5)$$

(d) For $V_D < -BV$,

$$I_D = -IS \cdot \left\{ \exp\left(\frac{-(BV + V_D)}{V_t}\right) - 1 + \frac{BV}{V_t} \right\} \quad (3.6)$$

For all of these equations, V_t is the thermal voltage which is defined as

$$V_t = \frac{k \cdot T}{q} \quad (3.7)$$

To insure convergence between regions (c) and (d), it is necessary that IBV is defined as

$$IBV \geq \frac{IS \cdot BV}{V_t} \quad (3.8)$$

A typical plot of I_D versus V_D generated by these equations is shown in Figure 49 where each of the regions (a) through (d) are indicated.

The Small–Signal AC Model

The ac model of the SPICE diode is derived from the linearized small–signal behavior of the internal diode D_1 shown in Figure 48. The circuit elements of this model include the junction capacitance CJ, the dynamic conductance GD, and the diffusion capacitance CD, all of which are bias dependent as are those corresponding elements of the ideal diode model.

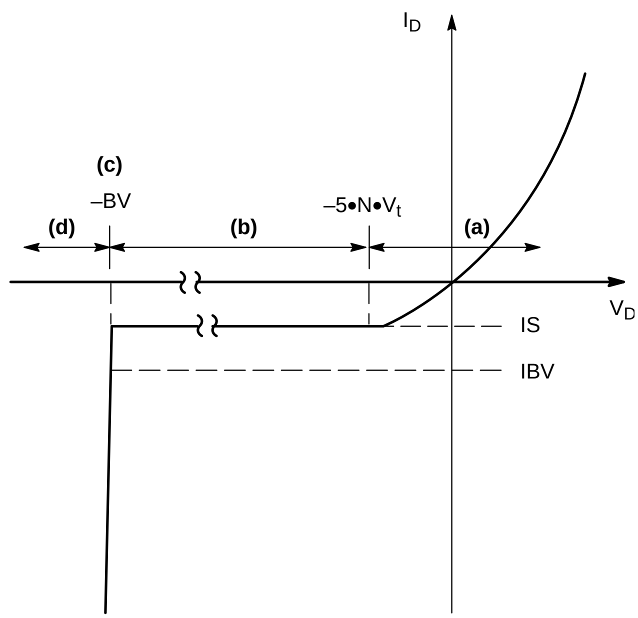


Figure 49. Large–scale I–V Characteristics of the SPICE Diode Model

The junction capacitance is modeled by the parameters CJO, VJ, M, and FC. The parameters CJO and VY are identical to the zero–bias junction capacitance $C_j(0)$ and the contact potential V_j described for an ideal diode. The parameter M is a grading exponent that is used to change the slope of the junction capacitance versus voltage (C–V) characteristics curve. For abrupt or step junctions, M is 0.5 while for linearly graded junctions, M is about 0.333. The parameter FC is used to model the capacitance under forward bias conditions.

The variables for the model are the junction capacitance CJ in farads and the internal diode voltage V_D which are related by

$$CJ = f(V_D) \quad (3.9)$$

where two regions of operation describe this function.

(a) For $V_D < FC \cdot VJ$,

$$CJ = CJO \cdot \left\{1 - \frac{V_D}{VJ}\right\}^{-M} \quad (3.10)$$

(b) For $V_D \geq FC \cdot VJ$,

$$CJ = \frac{CJO}{(1 - FC)^{(M+1)}} \cdot \left(1 - FC \cdot (M + 1) + \frac{M \cdot V_D}{VJ}\right) \quad (3.11)$$

To insure that the last equation remains well behaved, PC is restricted to values between zero and one; that is,

$$0 \leq FC < 1 \quad (3.12)$$

A typical C–V curve produced by these equations is shown in Figure 50 where each of the two regions are indicated.

The dynamic conductance is modeled by the slope of the I–V curve evaluated at a particular bias voltage. This slope is found from the voltage derivative of the current described in equations (3.3) through (3.6). The conductance GD in mhos of the internal diode D_1 as a function of the voltage V_D is derived from

$$GD = f(V_D) \quad (3.13)$$

where three regions of operation describe this functional relationship which involve the parameters IS and N.

(a) For $V_D \geq -5 \cdot N \cdot V_t$,

$$G_D = \left(\frac{IS}{N \cdot V_t}\right) \cdot \exp\left(\frac{V_D}{N \cdot V_t}\right) \quad (3.14)$$

(b) For $-BV < V_D < -5 \cdot N \cdot V_t$,

$$GD = -\frac{IS}{V_D} \quad (3.15)$$

(c) For $V_D \leq -BV$,

$$GD = 0 \quad (3.16)$$

The diffusion capacitance CD is modeled by the forward transit time parameter TT**,** and the parameters IS and N. Similar to an ideal diode, this capacitance in farads is voltage dependent and is derived from

$$CD = f(V_D) \quad (3.17)$$

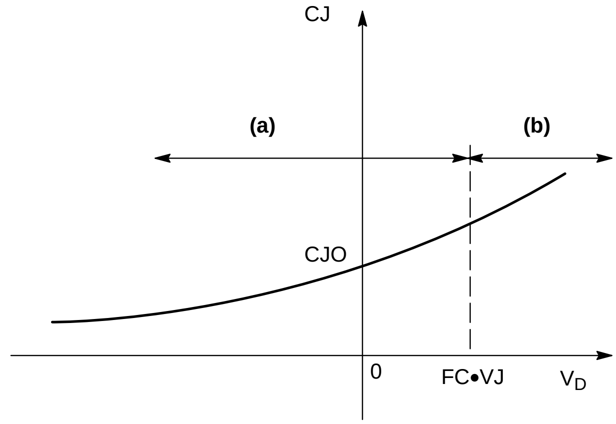


Figure 50. Junction C–V Characteristics of the SPICE Diode Model

where three regions of operation describe this functional relationship:

(a) For $V_D \geq -5 \cdot N \cdot V_t$

$$C_D = \left(\frac{TT \cdot IS}{N \cdot V_t}\right) \cdot \exp\left(\frac{V_D}{N \cdot V_t}\right) \quad (3.18)$$

(b) For $-BV < V_D < -5 \cdot N \cdot V_t$

$$CD = -\frac{TT \cdot IS}{V_D} \quad (3.19)$$

(c) For $V_D \leq -BV$

$$CD = 0 \quad (3.20)$$

The complete small–signal ac model of the SPICE diode is shown in Figure 51 where the resistance RS has been included with the elements defined above.

Temperature and Area Effects

Temperature behavior of the SPICE diode is modeled through certain temperature–dependent parameters. These parameters are IS, VJ, CJO, and EC. In the equations to follow, TNOM is the nominal or reference temperature having a default value of 27°C. This value can also be changed with the use of the .OPTIONS line. The variable T is the analysis temperature (in °C) which has a default value of 27°C if the TEMP line is omitted from the circuit file. If the .TEMP line is used for temperature analysis, T takes on the values given in this line. It is important to note that SPICE assumes all input data and model parameters have been specified at 27°C. Even though temperature is specified in the circuit file and in the .OPTIONS line in °C, it is converted by SPICE to °K for use in the equations.

For the saturation current IS, the parameters N, EG, and XTI are used to model its temperature dependence with the equation:

$$IS(T) = IS(TNOM) \cdot \left(\frac{T}{TNOM}\right)^{XTI1N} \cdot \exp\left\{\left[\frac{q \cdot EG}{N \cdot K}\right] \cdot \left[\frac{1}{TNOM} - \frac{1}{T}\right]\right\} \quad (3.21)$$

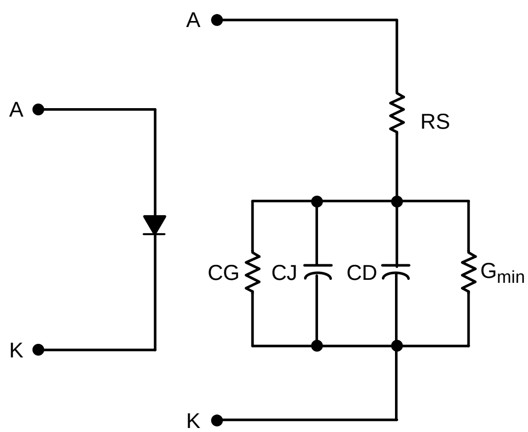


Figure 51. AC Small–signal SPICE Diode Model

where IS (TNOM) is the nominal value of the saturation current and IS(T) is the value evaluated at the analysis temperature T.

The contact potential VJ has the temperature–dependent function given below

$$VJ(T) = VJ(TNOM) \cdot \left(\frac{T}{TNOM}\right) + \frac{2 \cdot k \cdot T}{q} \cdot 1n\left(\frac{n_i(TNOM)}{n_i(T)}\right) \quad (3.22)$$

In this expression, VJ(TNOM) is the TNOM value of VJ, VJ(T) is the value evaluated at T, and $n_i(T)$ is the intrinsic carrier concentration (in cm^{-3}) of silicon which is also a function of temperature. That is,

$$n_i(T) = 1.45 \cdot 10^{10} \cdot \left(\frac{T}{TNOM}\right)^{1.5} \times \exp\left\{\left[\frac{q}{2 \cdot K}\right] \cdot \left[\frac{E_g(TNOM)}{TNOM} - \frac{E_g(T)}{T}\right]\right\} \quad (3.23)$$

where $E_g(T)$ is the temperature–dependent function for the energy gap (in eV) of silicon. From experimental results, this function is found to be

$$E_{g(T)} = 1.16 - \frac{7.02 \cdot 10^{-4} \cdot T^2}{T + 1108.0} \quad (3.24)$$

At 27°C, E_g calculates to be about 1.115 eV.

For the zero–bias junction capacitance CJO, the parameters M and VJ are used to model its temperature dependence with the equation:

$$CJO(T) = CJO(TNOM) \cdot \left\{1 + 4 \cdot 10^{-4} \cdot M \cdot (T - TNOM) + m \cdot \left[1 - \frac{VJ(T)}{VJ(TNOM)}\right]\right\} \quad (3.25)$$

The last temperature–dependent parameter is FC, the junction capacitance forward–bias coefficient. This parameter has the simple function given as:

$$FC(T) = FC(TNOM) \cdot \left(\frac{VJ(T)}{VJ(TNOM)}\right) \quad (3.26)$$

For the ideal diode, certain parameters are functions of the junction area. The SPICE program also provides area–dependency on these parameters through the use of the AREA factor in the diode description line in the circuit file. The parameters affected by the AREA factor are IS, RS, CJO and IBV which are modified by the following equations:

$$IS = AREA \cdot IS \quad (3.27)$$

$$RS = \frac{RS}{AREA} \quad (3.28)$$

$$CJO = AREA \cdot CJO \quad (3.29)$$

$$IBV = AREA \cdot IBV \quad (3.30)$$

where AREA has a default value of 1.

The Noise Model

Small–signal noise behavior of the SPICE diode is modeled by two noise current sources added to the small–signal ac circuit model as shown in Figure 52. The current source i_{RS} is responsible for modeling thermal noise generated by the resistance RS. The mean–squared value of thermal noise current (in A^2) generated by this source is expressed as

$$\overline{i_{RS}^2} = \frac{4 \cdot k \cdot T}{RS} \cdot \Delta f \quad (3.31)$$

where T is the temperature in °K and Δf is the noise bandwidth in Hz. The current source i_D, is responsible for modeling both shot and flicker noise (1/f noise) generated in the depletion region of the diode. The total mean–squared value of noise current (in A^2) generated by this source is expressed as

$$\overline{i_D^2} = 2 \cdot q \cdot I_D \cdot \Delta f + KF \cdot \frac{I_D}{f} \cdot \Delta f \quad (3.32)$$

where I_D is the dc diode current, f is the frequency at which the noise is measured, and AF and KF are SPICE flicker noise parameters.

The Diode Model versus a Real Diode

To see how well the equations for the SPICE diode model simulate the behavior of both ideal and real diodes, comparisons are made among the terminal characteristics of all three. These comparisons use standard characteristic curves which illustrate the behavior of a typical real diode (shown with solid curves), and those produced by the ideal and SPICE diode models (shown with dashed curves). For clarity, all curves are labeled to indicate their origin.

The current (logarithmic scale) versus voltage (linear scale) for a diode driven by a forward–bias dc voltage is plotted and shown in Figure 53. From this I–V plot,three distinct regions of operation are observed. In Region I, the device is operating under *extreme low–level injection* for bias voltages ranging from zero to about 100 mV, typically. For this mode of operation, the injected carriers passing through the depletion region are largely affected by the many generation and recombination (G–R) centers found near the metallurgical junction. The dominant effect of these G–R centers causes the increase in the diode current with respect to the voltage to be smaller than ideally predicted. In Region II, the diode is operating under *low–level injection.* The voltage range for this region is from about 100 mV to a value determined by $V_F(max)$ which is derived from

$$V_F(max) = V_t \cdot 1n\left(\frac{N_{low}^2}{10 \cdot n_i^2}\right) \qquad (3.32)$$

where N_{low} is the smaller of the impurity concentrations. Under this mode of operation, the G–R centers do not affect the injected carriers as much as in Region I, and the current increase over this voltage range is found to be close to that predicted for an ideal diode. Beyond the voltage $V_F(max)$, the diode operates under *high–level injection* as illustrated by Region III. Here the resistances of the neutral n and p–type regions produce ohmic voltage drops which tend to reduce the current increase. As these drops become more dominant, the diode current stops increasing exponentially and becomes more proportional to the bias voltage.

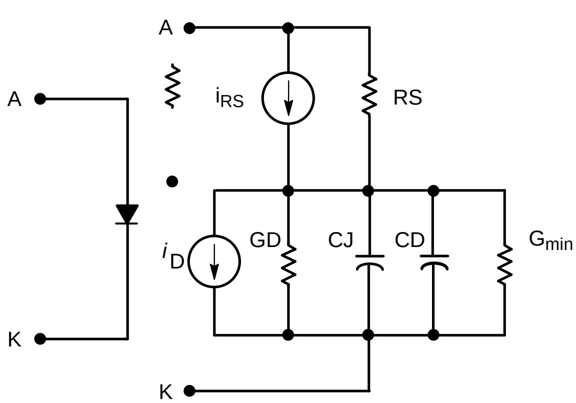


Figure 52. AC Small–signal SPICE Diode Model with Noise Sources

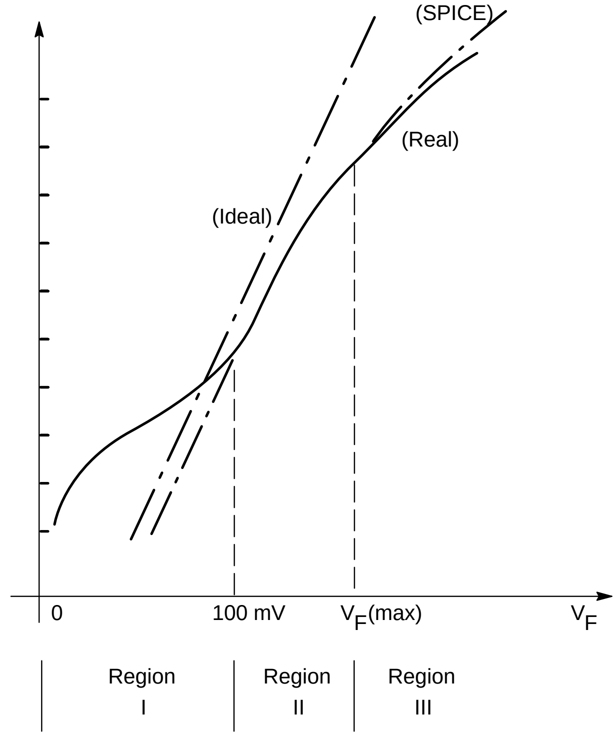


Figure 53. Forward–bias I–V Characteristics for Real, Ideal, and SPICE Diodes

Superimposed on this plot is the dashed curve representing the I–V characteristics of an ideal diode. It is clear that this model is fairly accurate in Region II, but falls short of predicting the behavior of Regions I and III. The second dashed curve is that of the SPICE diode which is generated by Equations 3.1 and 3.3. Compared to the real diode, the characteristics produced by this model illustrate significant accuracy in Regions II and III. However, like the ideal diode, the characteristics are less accurate in Region I.

In Figure 54, the linear–scaled I–V characteristics of a real diode are shown for both the forward and reverse–bias conditions. As the reverse–bias voltage increases toward the breakdown voltage, the diode current in the reverse direction exhibits a slight increase due to surface leakage effects. For real diodes, this current is referred to as the *reverse leakage current I_R.* The current at which breakdown occurs is called the *breakdown current I_{BV}.* The first dashed curve represents the I–V characteristics of an ideal diode. Under reverse–bias conditions, the reverse leakage current is main–mined by the saturation current I_S which is constant and not affected by the voltage. Even though the breakdown voltage can be calculated, the true breakdown behavior is not predicted by the ideal diode model equations.

The second dashed curve represents the I–V characteristics of the SPICE diode as generated by Equations 3.3 through 3.6. For this model, the reverse leakage current is maintained by the saturation current parameter IS, which is also constant and unaffected by the voltage. However, the behavior of the model at the breakdown voltage is fairly close to that of a real diode.

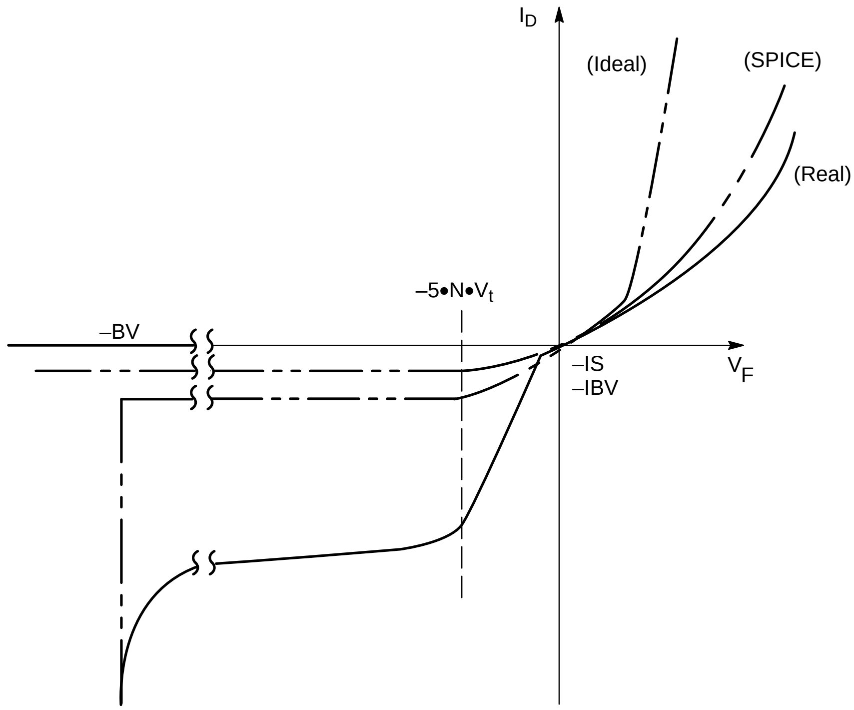


Figure 54. Forward and Reverse–bias I–V Characteristics for the Real, Ideal, and SPICE Diodes

Junction capacitance versus voltage characteristic curves are shown in Figure 55. Under forward–bias, the C–V characteristics for a real diode are well behaved for voltage values close to the contact potential V_j. For an ideal diode, however, the capacitance calculated near V_j tends to approach unrealistic values. In the reverse–bias region, the results of this equation show close similarity to the capacitance of a real diode.

The results of the SPICE diode junction capacitance equations show a very accurate similarity to the capacitance of a real diode, especially in the reverse–bias region. This is due in part to the. parameter M in Equation 3.10 which allows the slope of the C–V curve to vary, lathe forward–bias region, the parameter PC is used in Equation 3.11 to produce a straight–line approximation to the C–V curve for voltages beyond FC•VJ.

A plot of diffusion capacitance as a function of forward–bias diode current is shown in Figure 56. The diffusion capacitance behavior of an ideal diode illustrates a proportional increase in capacitance with current since the forward transit time T is assumed to be constant. This is the case for the SPICE diode since the corresponding parameter TT is also constant. For a real diode, however, τ_T is known to increase with current at high current levels due to current crowding effects. Thus, the diffusion capacitance typically deviates from predicted behavior as illustrated.

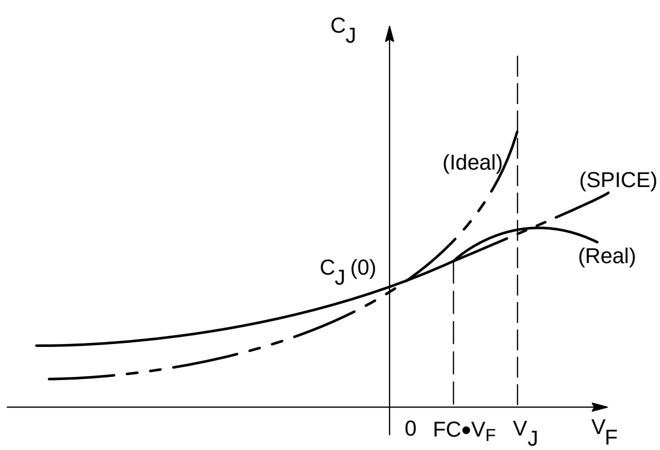


Figure 55. Forward and Reverse–bias Junction C–V Characteristics for the Real, Ideal, and SPICE Diodes

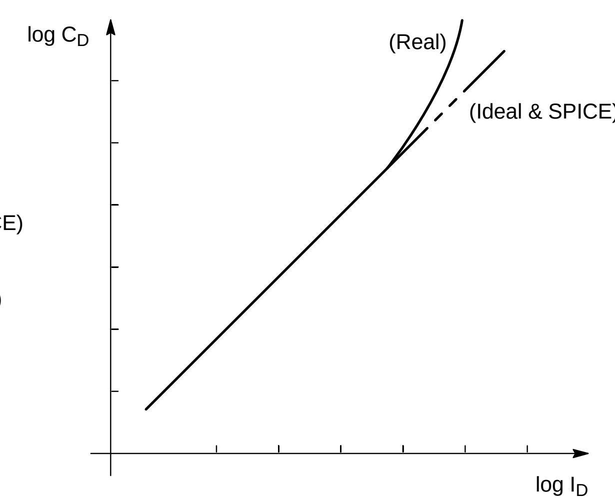


Figure 56. Diffusion Capacitance Characteristics for the Real, Ideal, and SPICE Diodes

Figure 57 is a plot of the temperature coefficient (TCV) as a function of forward–bias current. The TCV for a real diode typically has negative values over low current ranges but increases to positive values at higher currents. This is due mainly to the positive temperature coefficient of the bulk resistance which dominates the diode voltage at these levels. The TCV of an ideal diode has negative values over a large range of current as shown. This is also true for the SPICE diode since the resistance parameter RS has no temperature coefficient.

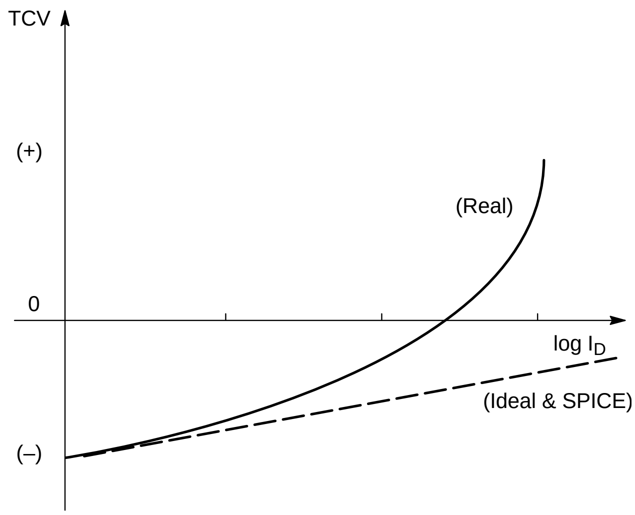


Figure 57. Voltage Temperature Coefficient (TCV) Characteristics for the Real, Ideal, and SPICE Diodes

SPICE Diode Model Parameter Limitations and Restrictions

Based on these comparisons, it is obvious that the SPICE diode model is capable of performing a fairly accurate job in simulating the behavior of a real pn junction diode. It is just as obvious, however, that the model is also limited in its range of accurate simulation due in part to limitations of its parameters. The following statements provide a review of the limitations on the model parameters and how they affect the simulation results. For some of the parameters, suggested value restrictions are given in order to insure convergence.

(a) The saturation current IS is constant and not a function of the reverse–bias voltage. Therefore, the simulated reverse leakage current remains constant over the reverse–bias region up to the breakdown voltage.

(b) The ohmic resistance RS is constant and not a function of current or voltage. The resistance of the bulk neutral semiconductor regions of a real diode actually increases as current increases for high–level injection.

(c) RS has no temperature coefficient. The temperature coefficient of the resistance of the bulk neutral semiconductor regions of a real diode is actually positive, which produces a positive TCV at high–level injection.

(d) The emission coefficient N is constant, and cannot model the change in the slope of the I–V characteristics between the extreme low–level and low–level injection regions. For convergence purposes, N must be greater than 0.01.

(e) The forward transit time TT is constant and not a function of current or voltage. As such, the diffusion capacitance CD will increase proportionally with current and the simulated reverse recovery time t_{rr} will be constant over current.

(f) The temperature dependence of the zero–bias junction capacitance CJO is consistent with that of silicon only. In Equation 3.25, the coefficient of thermal expansion is that of silicon material.

(g) From Equations 3.22 to 3.24, the TNOM value of the contact potential VJ must be greater than 0.4 V to insure convergence for temperature analysis up to 200°C. For a larger value of the analysis temperature, these equations can be used to determine the minimum value of VJ.

(h) The temperature dependencies of the contact potential VJ, the intrinsic cater concentration ni, and the energy gap EG given in Equations 3.22 to 3.24, respectively, are consistent with that of silicon material only.

(i) For convergence, the forward bias coefficient PC is restricted to values between zero and one as indicated in Equation 3.12; that is,

$$0 \le FC < 1.0$$

(j) The reverse breakdown voltage BV has a default value of infinite. A specified value of zero for BV is interpreted by SPICE to mean infinite.

(k) The reverse breakdown characteristics of a real diode tend to be “soft”; that is, the reverse leakage current gradually increases toward the breakdown current as the reverse–bias voltage increases. For the SPICE model, the reverse leakage current is modeled by the parameter IS which is constant out to the breakdown voltage. This produces a “hard” breakdown characteristic for the model.

(l) The current at the breakdown voltage 1EV is dependent upon IS and BV. For convergence, IBV is restricted to values determined by

$$IBV \ge \frac{IS(T) \cdot BV}{V_{t(T)}} \qquad (3.34)$$

where the saturation current IS and the thermal voltage V_t must be calculated at the largest value of simulation temperature by Equations 3.21 and 3.7, respectively.

(m) IBV is often not consistent with the breakdown current IBV of a real diode at the specified breakdown voltage.

SPICE Diode Model Parameter Extraction Methods

The methods given in this section for the extraction of the SPICE diode model parameters are based on data acquired from a real device, Table 6 lists the necessary data required for the extraction of certain parameters. This data may be taken from actual measurements or from characteristic plots available on most data sheets. For data sheet information, it is assumed that the plots represent a typical device and that the SPICE model is valid for these devices. The tools needed for the extraction methods include a suitable scientific calculator (programmable, if possible) which is capable of performing statistical calculations, and an understanding of linear and nonlinear least–squares curve fitting methods.

Table 6. Diode Data Requirements for SPICE Model Parameter Extraction

Data Required	Parameters Extracted
Forward dc characteristics; forward diode current (I_D) versus forward diode voltage (V_F).	IS, RS, N
Junction capacitance characteristics; reverse–bias junction capacitance (C_j) versus reverse bias voltage (V_R).	CJO, VJ, M
Reverse recovery time (t_{rr}) versus forward diode current (I_F).	TT
Forward voltage (VF) temperature coefficient versus forward current (ID), or reverse current (IR) versus temperature.	EG, XTI
Breakdown voltage characteristics.	BV, IBV
Noise measurements (parameters usually set to default values).	KF, AF
Other parameters for which the defaults are assumed: FC = 0.5 (typical)	FC

For some parameters, more than one method of extraction may be given. To determine which method is *best* in providing a set of parameters yielding the most accurate simulation, it is necessary to examine the differences between actual device data and the results provided by the model equations when using these parameters. By defining these differences as *errors,* a single quantity can be calculated for each method that can be used as a basis of comparison. This quantity is called the *error function* E_2 which is the sum of the squares of the magnitudes of the per unit or normalized errors between actual device data and the corresponding data generated by the model. This function is represented by the equation given below

$$E_2 = |\varepsilon_1|^2 + |\varepsilon_2|^2 \ldots + |\varepsilon_n|^2 = \Sigma|\varepsilon_i|^2 \qquad (3.35)$$

where ε_i, is the normalized error of the ith data point and *n* is the number of data points [7], For most modeling applications, the method which produces the smallest value of E_2 is therefore judged the best method and the parameters extracted with this method provide the most accurate results.

Forward DC Characteristics (IS, RS, N)

There are three methods for the extraction of the parameters IS, RS, and N which model the large–signal dc behavior in the forward–bias region. The first method computes these parameters from three points taken from the forward–bias I–V curve and is appropriately called the *three–point I–V method.* The second method uses a fixed value of RS (which can be taken from the results of the three–point I–V method) and performs a linear regression data fit over the *I–V* curve to extract IS and N. The third method uses a fixed value of IS and performs a nonlinear data fit to extract N and RS, This method is useful if the reverse leakage current I_R is specified and is to be modeled by IS.

1. *Method 1 (Three–point 1–V method).* To use this method, it is necessary to have a plot of the dc forward–bias I–V curve where the current axis is logarithmically scaled and the voltage axis is linearly scaled as shown in Figure 58. Estimated values for the parameters can be found from the steps outlined below.

(i) Select three data points from the I–V curve shown as points 1, 2, and 3.

(ii) Using the values of current and voltage corresponding to these points as indicated, calculate values for the parameters RS, N, and IS from the following equations

$$RS = \frac{(V_{F2} - V_{F1}) + (V_{F1} - V_{F3}) \cdot \left\{ \dfrac{1n\left(\frac{I_{D1}}{I_{D2}}\right)}{1n\left(\frac{I_{D1}}{I_{D2}}\right)} \right\}}{(I_{D2} - I_{D1}) + (I_{D1} - I_{D3}) \cdot \left\{ \dfrac{1n\left(\frac{I_{D1}}{I_{D2}}\right)}{1n\left(\frac{I_{D1}}{I_{D2}}\right)} \right\}} \qquad (3.36)$$

$$N = \frac{(V_{F1} - V_{F2}) + RS \cdot (I_{D2} - I_{D1})}{V_t \cdot 1n\left(\frac{I_{D1}}{I_{D2}}\right)} \qquad (3.37)$$

$$IS = \frac{I_{D1}}{\exp\left(\frac{V_{F1} - RS \cdot I_{D1}}{N \cdot V_t}\right) - 1} \qquad (3.38)$$

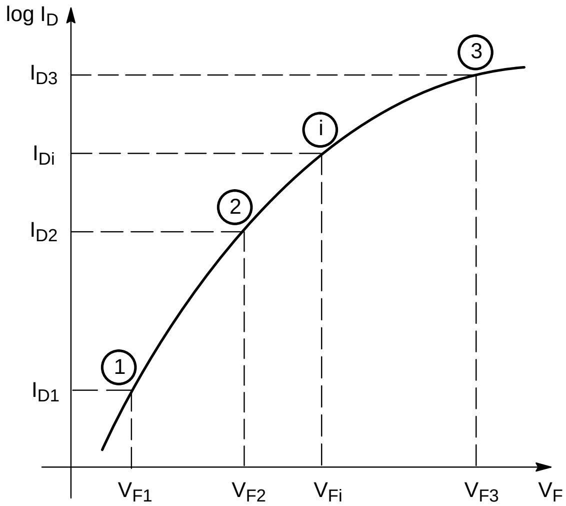


Figure 58. Forward–bias DC Data for the Three–point I–V Method

2. *Method 2 (Linear regression with fixed RS).* With this method, a linear regression data fit is performed on the data covering the full range of the I–V curve to an equation of a straight line which represents the diode dc characteristics. This equation can be derived by combining Equations 3.1 and 3.3 to produce

$$I_{Di} = IS \cdot \left\{ \exp\left(\frac{V_{Fi} - RS \cdot I_{Di}}{N \cdot V_t}\right) - 1 \right\} \qquad (3.39)$$

where I_{Di} and V_{Fi} represent the ith data point taken from *n* data points on the I–V curve (shown in Figure 59), and RS is a known value of the ohmic resistance parameter. Assuming that the exponential term is sufficiently large, this equation can be approximated by

$$1n(I_{Di}) = 1n(IS) + \frac{1}{N} \cdot \left(\frac{V_{Fi} - RS \cdot I_{Di}}{V_t}\right) \qquad (3.41)$$

Taking the natural logarithm of both sides yields

$$y_i = b + m \cdot x_i \qquad (3.42)$$

which is the equation of a straight line expressed as

$$I_{Di} \cong IS \cdot \exp\left(\frac{V_{Fi} - RS \cdot I_{Di}}{N \cdot V_t}\right) \qquad (3.40)$$

where:

$$y_i = 1n(I_{Di}) \qquad (3.43)$$

$$x_i = \left(\frac{V_{Fi} - RS \cdot I_{Di}}{V_t}\right) \qquad (3.44)$$

$$b = 1n(IS) \qquad (3.45)$$

and

$$m = \frac{1}{N} \qquad (3.46)$$

The constants b and m are the y–axis intercept and slope, respectively, of the straight line represented by Equation 3.42, Values for these constants can be found by solving the matrix equation

$$\begin{bmatrix} n & \Sigma(x_i) \\ \Sigma(x_i) & \Sigma(x_i)^2 \end{bmatrix} \cdot \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} \Sigma(y_i) \\ \Sigma(x_i) \cdot (y_i) \end{bmatrix} \qquad (3.47)$$

where the summations are taken over the n data points. The calculated values of b and m are then used to find IS and N from Equations 3.45 and 3.46 where

$$IS = \exp(b) \qquad (3.48)$$

and

$$N = \frac{1}{m} \qquad (3.49)$$

It is important to note that many scientific calculators have built–in statistical functions capable of performing linear regression calculations. For the type of equations used in this method, these functions can be used to extract the parameters IS and N directly. This, of course, eliminates the need to generate and solve Equation 3.47. The steps involved in this method are outlined below.

(i) Select a value of RS which maybe the value calculated from Equation 3.36 in Method 1.

(ii) Perform a linear regression data fit on *n* data points taken from the I–V curve to Equation 3.42 where the yi and xi data values are calculated from Equations 3.43 and 3.44. This data fit will give values for *b* and *m.*

(iii) Calculate the values of IS and N from Equations 3.48 and 3.49, respectively.

3. *Method 3 (Nonlinear curve fit with fixed IS).* The techniques used in this method are similar to those of Method 2 except that IS is held fixed, and RS and N are extracted. This method is particularly useful when IS must model a known value of the reverse leakage current. The equation for the curve fit is derived from Equation 3.39 where voltage is expressed as the dependent variable; that is

$$V_{Fi} = N \cdot V_t 1n\left(\frac{I_{Di}}{IS} + 1\right) + RS \cdot I_{Di} \qquad (3.50)$$

where I_{Di} and V_{Fi} again represent the ith data point taken from n data points on the I–V curve, and IS is a known value of the saturation current parameter. This equation may be expressed as a linear combination of two functions of the current in the form of

$$y_i = a_1 \cdot f_1(I_{Di}) + a_2 \cdot f_2(I_{Di}) \qquad (3.51)$$

where:

$$y_i = V_{Fi} \quad (3.52)$$

$$f_1(I_{Di}) = V_t \cdot 1n\left(\frac{I_{Di}}{IS} + 1\right) \quad (3.53)$$

$$f_2(I_{Di}) = I_{Di} \quad (3.54)$$

$$a_1 = N \quad (3.55)$$

and

$$a_2 = RS \quad (3.56)$$

The constants a_1 and a_2 are found from the solution of a matrix equation similar in appearance to that of Equation 3.47 where

$$\begin{bmatrix} \Sigma[f_1(I_{Di})]^2 & \Sigma[f_1(I_{Di})f_2(I_{Di})] \\ \Sigma[f_1(I_{Di})f_2(I_{Di})] & \Sigma[f_1(I_{Di})]^2 \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \Sigma[y_1 \cdot f_1(I_{Di})] \\ \Sigma[y_1 \cdot f_2(I_{Di})] \end{bmatrix} \quad (3.57)$$

The summations for the elements of this matrix equation are again taken over the n data points. Values of a_1 and a_2 found from this equation are then used to calculate N and RS where:

$$N = a_1 \quad (3.58)$$

and

$$RS = a_2 \quad (3.59)$$

The details explaining more on the techniques of this method may be found in [8], The steps involved in this method are outlined below.

(i) Select a value of IS which may be that of the reverse leakage current at a value of specified reverse–bias voltage.

(ii) Generate the matrix Equation 3.57 where y_i and the functions $f_1(I_{Di})$ and $f_2(I_{Di})$ are calculated from Equations 3.52 through 3.54 for the n data points taken from the I–V curve.

(iii) Solve this matrix equation for the constants a_1 and a_2, and calculate the values of N and RS from Equations 3.58 and 3.59, respectively.

Junction Capacitance Characteristics (CEO, VJ, M, FC)

There are two methods for the extraction of the parameters CJO, VJ, M and FC which model the behavior of the junction capacitance. The first method computes these parameters from three points taken from the reverse–bias C–V curve and is appropriately called the *three–point C–V method*. The second method uses a fixed value of VJ (which can be taken from the results of the three–point C–V method) and performs a linear regression data fit over the C–V curve to extract CJO and M. In both methods, the parameter FC is set to the default value of 0.5 since forward–bias capacitance information is rarely presented on most diode data sheets.

1. *Method 1 (Three–point C–V method)*, To use this method, it is necessary to have a plot of the reverse–bias junction capacitance curve (C_j versus V_R) where each axis is logarithmically scaled as shown in Figure 59.

(i) Select a data point at the lower end of the C–V curve shown as point 1. The voltage at this point should be less than the typical ideal value of VJ (that is, 0.8 volt to 1.0 volt).

(ii) Select two data points at the upper end of the curve shown as points 2 and 3. The voltages at these points should be much greater than the typical ideal value of VJ.

(iii) Using the values of capacitance and voltage corresponding to these points as indicated, calculate values for the parameters M, VJ, CJO, and FC from the following equations

$$M = \frac{1n\left(\frac{C_{j2}}{C_{j3}}\right)}{1n\left(\frac{V_{R3}}{V_{R2}}\right)} \quad (3.60)$$

$$k_1 = \left(\frac{C_{j1}}{C_{j2}}\right)^{1/M} \text{ (a constant)} \quad (3.61)$$

$$VJ = \frac{k_1 \cdot V_{R1} - V_{R2}}{1 - k_1} \quad (3.62)$$

$$CJO = C_{j1} \cdot \left(1 + \frac{V_{R1}}{VJ}\right)^M \quad (3.63)$$

$$FC = 0.5 \quad (3.64)$$

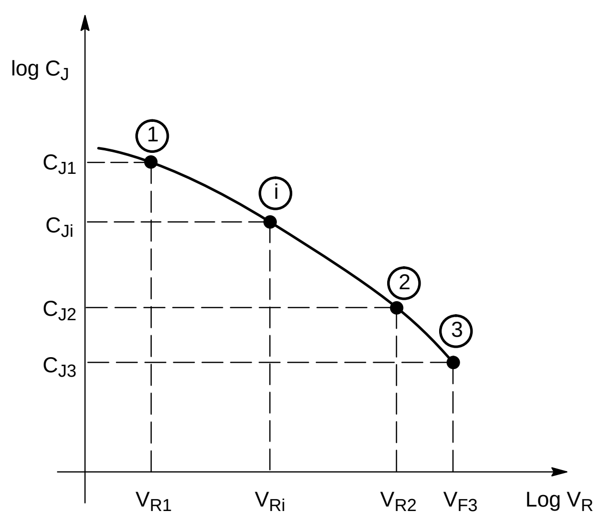


Figure 59. Reverse–bias Junction Capacitance Data for the Three–point C–V Method

2. *Method* 2 *(Linear regression with fixed VJ).* With this method, a linear regression data fit is performed on the data covering the full range of the reverse–bias C–V curve to an equation of a straight line representing the capacitance characteristics in this region. For reverse–bias voltages, Equation 3.10 can be expressed as

$$C_{ji} = CJO \cdot \left(1 + \frac{V_{Ri}}{VJ}\right)^{-M} \quad (3.65)$$

where C_{ji} and V_{Ri} represent the ith data point taken from n data points on the C–V curve (shown in Figure 59), and VJ is a known value of the contact potential parameter. Taking the natural logarithm of both sides yields

$$1n(C_{ji}) = 1n(CJO) - M \cdot 1n\left(1 + \frac{V_{Ri}}{VJ}\right) \quad (3.66)$$

which is the equation of a straight line expressed as

$$Y_i = b + m \cdot x_i \quad (3.67)$$

where:

$$y_i = 1n(C_{ji}) \quad (3.68)$$

$$x_i = 1n\left(1 + \frac{V_{Ri}}{VJ}\right) \quad (3.69)$$

$$b = 1n(CJO) \quad (3.70)$$

and

$$m = -M \quad (3.71)$$

The constants b and m are the y–axis intercept and slope, respectively, of the straight line represented by Equation 3.67 Values for these constants can be found by solving a matrix equation similar to that of equation 3.47 or by using the linear regression function on the calculator. The parameters CJO and M are then calculated from

$$CJO = \exp(b) \quad (3.72)$$

and

$$M = -m \quad (3.73)$$

The steps for this method are outlined below.

(i) Select a value of VJ which may be the value calculated from Equation 3.62 in Method 1.

(ii) Perform a linear regression data fit on n data points taken from the C–V curve to Equation 3.67 where the y_i and x_i data values are calculated from Equations 3.68 and 3.69. This data fit will give values for b and m.

(iii) Calculate the values of CJO and M from Equations 3.72 and 3.73, respectively, and the value of FC from Equation 3.64.

Reverse Recovery Time Characteristics (TT)

For the extraction of the forward transit time parameter TT, the results of reverse recovery time (t_{rr}) measurements derived by Leinfelder are employed [9]. The common JEDEC test circuit for measuring t_{rr} shown in Figure 60 while the idealized time–domain waveform of the diode current resulting from this circuit is shown in Figure 61. The reverse recovery time t_{rr} is commonly measured between the time that the current (previously forward biased at I_F) passes through zero going negatively and the time that the reverse current recovers to a value which is less than 10% of the peak reverse current I_{RM} [10]. From Figure 61, t_{rr} is shown to consist of t_a, which is the time for the diffusion charge supporting the reverse current to reduce to zero, and t_b which is the time for the depletion charge supporting the forward voltage to also reduce to zero. At the end of t_{rr} the diode is turned off and cannot sustain the reverse current. The total charge depleted from the diode during the reverse recovery time is called the *reverse recovery charge* Q_{rr} which is the sum of the charges Q_a and Q_b represented by the areas under the waveform during times t_a and t_b, respectively. By assuming that Q_{rr} is dominated by the charge Q_a so that the reverse recovery time t_{rr}, is approximately equal to t_a, the following equations can be generated from the information shown on Figure 61. For the current fall–time t_f:

$$t_f = \frac{I_F}{\left(\frac{di}{dt}\right)} \quad (3.74)$$

where I_F is the forward–bias diode current in amps and. di/dt is the slew–rate in amp/second of the current set by the circuit. The expression for t_{rr} is:

$$t_{rr} \cong t_a = \frac{I_{RM}}{\left(\frac{di}{dt}\right)} \quad (3.75)$$

where I_{RM} is the peak reverse current. Assuming that the reverse recovery charge Q_{rr} is approximately equal to the charge Q_a, the area under the triangle corresponding to this charge is computed from

$$Q_{rr} \cong Q_a = \frac{t_a \cdot I_{RM}}{2} = \frac{t_a^2}{2} \cdot \left(\frac{di}{dt}\right) = \frac{t_{rr}^2}{2} \cdot \left(\frac{di}{dt}\right) \quad (3.76)$$

From Leinfelder's work, an expression for Q_{rr} was derived from curve–fitting methods and was found to be accurate for simulating t_{rr}. This expression is given as

$$Q_{rr} = I_F \cdot TT \cdot \exp\left(-\sqrt{\frac{t_f}{TT}}\right)$$

$$= I_F \cdot TT \cdot \exp\left[-n\sqrt{\frac{I_F}{TT \cdot \left(\frac{di}{dt}\right)}}\right] \quad (3.77)$$

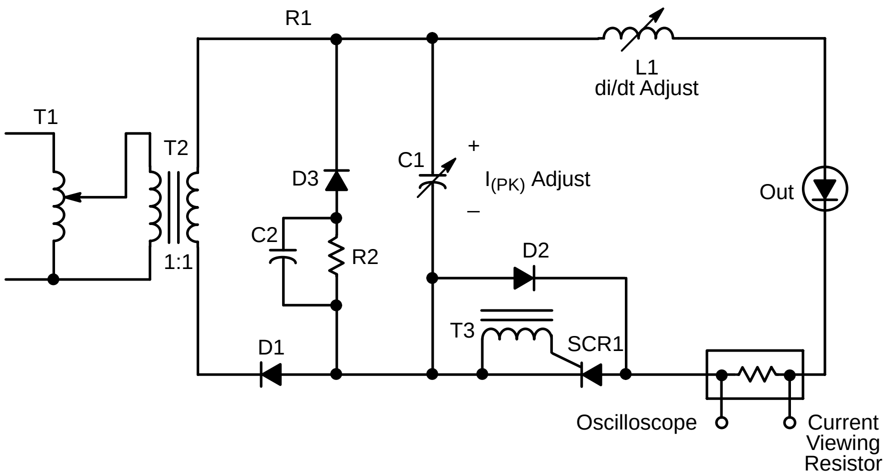


Figure 60. JEDEC Reverse Recovery Time Test Circuit

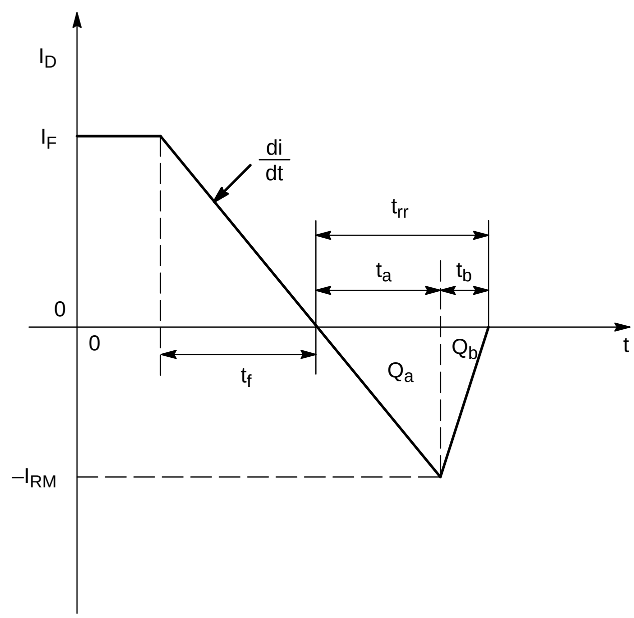


Figure 61. Idealized Time–domain Waveform for the Diode Current from the Test Circuit of Figure 60

where TT is, of course, the forward transit time parameter. By setting Equations 3.76 and 3.77 equal to each other, a single expression involving TT is derived

$$\frac{t_{rr}^2}{2} \cdot \left(\frac{di}{dt}\right) = I_F \cdot TT \cdot \exp\left(-n\sqrt{\frac{I_F}{TT \cdot \left(\frac{di}{dt}\right)}}\right) \quad (3.78)$$

Thus, by knowing t_{rr} I_F, and the current slew–rate di/dt, Equation 3.78 can be solved iteratively to find TT. Numerical algorithms such as *Newton's method* can be used on this equation which is easily implemented on a programmable calculator [7].

The steps for extracting the parameter TT are outlined below.

(i) From the device data sheet, determine values for the reverse recovery time t_{rr} the forward–bias diode current I_F, and the slew–rate of the current waveform di/dt.

(ii) Generate a function of the forward transit time TT where:

$$f(TT) = \frac{t_{rr}^2}{2} \cdot \left(\frac{di}{dt}\right) = I_F \cdot TT \cdot \exp\left(-n\sqrt{\frac{I_F}{TT \cdot \left(\frac{di}{dt}\right)}}\right) \quad (3.79)$$

(iii) Use Newton's method to solve Equation 3.79 for the value of TT that will set f(TT) to zero.

Temperature Characteristics (EG, XTI)

There are three methods for the extraction of the parameters EG and XTI. In each of these methods, only one temperature–dependent expression is used where EG is selected and XTI is calculated. The first method uses the small–scale expression for the voltage temperature coefficient TCV found from the derivative of the voltage V_F with respect to temperature. The second method uses the large–scale expression for the TCV which is derived from the change of the diode forward voltage over large temperature changes. The third method uses the expression for the temperature dependent saturation current IS given in Equation 3.21 to model the temperature behavior of the reverse leakage current I_R. In all of the equations used in these methods, temperature values must be convened to degrees Kelvin (°K).

1. *Method 1 (Small–scale TCV method).* From Equations 3.1 and 3.3, the expression for the diode forward voltage as a function of temperature is derived as:

$$V_F(T) = N \cdot V_t(T) \cdot \mathrm{1n}\left(\frac{I_D}{IS(T)} + 1\right) + RS \cdot I_D \quad (3.80)$$

Using the temperature–dependent expression for the saturation current IS given in Equation 3.21, the small–scale or derivative form of the voltage temperature coefficient TCV is derived and shown below

$$TCV = \frac{dV_F(T)}{dT} \bigg|_{I_{D,T}} = \frac{\left\{N \cdot V_t(T) \cdot \mathrm{1n}\left(\frac{I_D}{IS(T)} + 1\right) - (EG + XTI \cdot V_t(T)\right\}}{T} \quad (3.81)$$

where I_D is the forward current and T is the temperature at which the TCV is measured, usually 27°C. This equation is now used to find EG and XTI with the steps outlined below.

(i) From the data sheet, determine the value of the TCV, and the values of the forward current I_D, and the temperature T corresponding to the TCV.

(ii) Select a suitable value for the energy gap EG which should correspond to the type of material used to process the diode (for example, 1.11 eV for silicon).

(iii) Solve for the value of the parameter XTI from Equation 3.81 where

$$XTI = \left(\frac{1}{V_t(T)}\right) \cdot \left\{N \cdot V_t(T) \cdot \mathrm{1n}\left(\frac{I_D}{IS(T)} + 1\right) - (EG + T \cdot TCV)\right\} \quad (3.82)$$

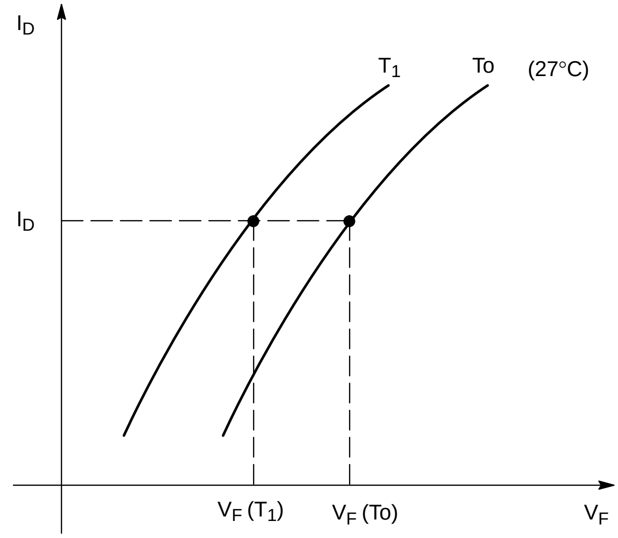


Figure 62. Typical Forward–bias Diode Voltage Temperature Characteristics

2. *Method 2 (Large–scale TCV method).* Again from Equation 3.80, the large–scale form of the voltage temperature coefficient is derived and shown below

$$TCV = \frac{\Delta V_F(T)}{\Delta T} = \frac{V_F(T_1) - V_F(T_0)}{T_1 - T_0} \bigg|_{I_D} = \left(\frac{1}{T_0}\right) \cdot \left\{N \cdot V_t(T_o) \cdot \mathrm{1n}\left(\frac{I_D}{IS(T_o)}\right) - \left[EG + XTI \cdot V_t(T_o) \cdot \left(\frac{T_1 \cdot \mathrm{1n}\left(\frac{T_1}{T_o}\right)}{T_1 - T_o}\right)\right]\right\} \quad (3.83)$$

where T_1 is a temperature read from the data sheet, T_o is room temperature of 27°C, $V_F(T_1)$ and $V_F(T_o)$ are the diode voltages at these temperatures, and I_D is the forward current, all of which are shown on the typical I–V curve of Figure 62. This equation and data are used to find EG and XTI with the steps outlined below.

(i) From the data sheet I–V curve plotted for at least two temperatures, calculate the large–scale TCV at the current I_D from the expression

$$TCV = \frac{V_F(T_1) - V_F(T_o)}{T_1 - T_o} \bigg|_{I_D} \quad (3.84)$$

(ii) Select a suitable value for the energy gap EG which should correspond to the type of material used to process the diode (for example, 1.11 eV for silicon).

(iii) Solve for the value of the parameter XTI from Equation 3.83 where

$$CTI = \frac{\left\{N \cdot V_t(T_o) \cdot \mathrm{1n}\left(\frac{I_D}{IS(T_o)}\right) - (EG + T_o \cdot TCV)\right\}}{V_t(T_o) \cdot \left(\frac{T_1 \cdot \mathrm{1n}\left(\frac{T_1}{T_o}\right)}{T_1 - T_o}\right)} \quad (3.85)$$

3. *Method 3 (Reverse leakage current method).* Assuming that the temperature behavior of the diode reverse leakage current I_R can be modeled by that of the saturation current parameter IS, Equation 3.21 can be used directly to extract the parameters EG and XTI. From this equation, the temperature–dependence of I_R can be written as

$$I_R(T_1) = I_R(T_o) \cdot \left(\frac{T_1}{T_o}\right)^{XTI1N} \cdot \exp\left\{\left(\frac{q \cdot EG}{N \cdot K}\right) \cdot \left(\frac{1}{T_o} - \frac{1}{T_1}\right)\right\} \quad (3.86)$$

where T_1 is a temperature read from the data sheet, T_o is room temperature of 27°C, and $I_R(T_1)$ and $I_R(T_o)$ are the leakage currents at these temperatures, all of which are shown on the typical I_R versus temperature curve of Figure 63. Equation 3.86 and this data are used to find EG and XTI with the steps outlined below.

(i) From the data sheet reverse leakage current versus temperature (I_R versus T) curve, select values of the leakage current at the temperatures T_o(27°C) and T_1 that is, $I_R(T_o)$ and $I_R(T_1)$.

(ii) Select a suitable value for the energy gap EG which should correspond to the type of material used to process the diode (for example 1.11 eV for silicon).

(iii) solve for the value of the parameter XTI from Equation 3.86 where

$$XTI = \frac{\left\{N \cdot 1n\left(\frac{I_R(T_1)}{I_R(T_o)}\right) - \frac{q \cdot EG}{k} \cdot \left(\frac{1}{T_o} - \frac{1}{T_1}\right)\right\}}{1n\left(\frac{T_1}{T_o}\right)} \quad (3.87)$$

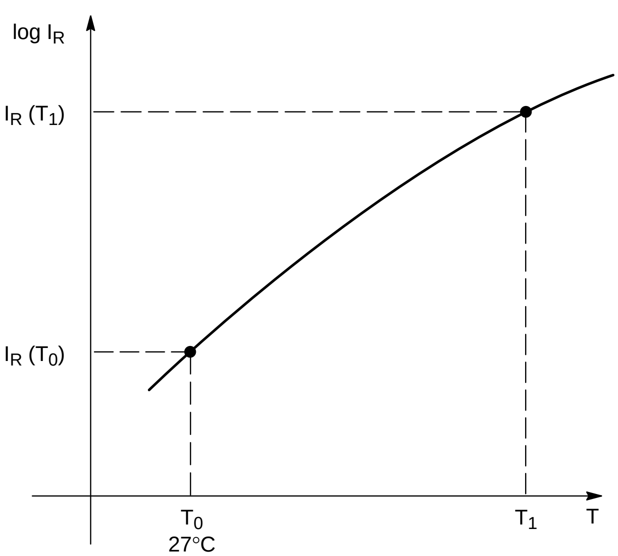


Figure 63. Typical Reverse Leakage Current Temperature Characteristics

Reverse Breakdown Characteristics (BV, IBV)

The steps for extracting the parameters BV and IBV are outlined below.

(i) From the device data sheet, determine the value of the maximum dc blocking voltage V_R(max).

(ii) Multiply this voltage by a constant k_{BV} having a value ranging from about 1.3 to 2.0. This will account for the amount of manufacturers guard–band placed on the breakdown voltage in order to insure the maximum voltage rating. This opera–don produces the value for the breakdown voltage parameter BV where

$$BV = k_{BV} \cdot V_R(max) \quad (3.88)$$

(iii) Calculate the value IBV from

$$IBV = (1.1) \cdot \frac{IS(T_{max}) \cdot BV}{V_t(T_{max})} \quad (3.89)$$

where the 1.1 multiplier insures the inequality in Equation 3.8 and T_{max} is the maximum simulation temperature.

Example Parameter Extractions for the MURHB4OCT Diode

As an example of the methods presented in the last section, the SPICE model parameters will be extracted for the ON Semiconductor MURH84OCT rectifier with the aid of a programmable scientific calculator. Pertinent characteristic curves taken from the device data sheet will be used to provide the necessary data for the extraction. In the processes that are to follow, particular attention will be placed not only on these curves but also on certain tabular information regarding maximum ratings and electrical characteristics. It is important that this data is included in the determination of the parameters since the model must simulate as closely as possible all characteristics of the diode.

It is industry practice to use 25°C as room temperature. For the extraction process, the SPICE default of 27°C will be used instead and it will be assumed that the 2°C difference will not produce any significant errors. At 27°C, the value for the thermal voltage V_t which is used in several of the extraction methods is calculated from Equation 3.7 where

$$V_t = 25.85562 \text{ mV} \quad (3.90)$$

For consistency, the order of parameter extraction will follow that of the last section which begins with the forward dc characteristics.

Forward DC Characteristics (IS, RS, N)

The forward–bias I–V curve for the MURH840CT rectifier is shown in Figure 64. Data from this curve will be used in each of the three methods for extracting the parameters IS, RS. and N.

1. *Method 1(Three–point I–V method)*. As shown in Figure 64, three points are selected on the I–V curve at 25°C as points 1 through 3. The current and voltage values corresponding to these points are listed in Table 7. From Equations 3.36 through 3.38, the following parameters are computed from this data for 27°C:

$$RS = 40.20590\ m\Omega \quad (3.91)$$

$$N = 2.854546 \quad (3.92)$$

and

$$IS = 3.255398\ \mu A \quad (3.93)$$

2. *Method 2 (Linear regression with fixed RS)*. From Figure 64, the I–V data given in Table 8 is obtained. These 10 I–V data points are transformed into x–y data points with Equations 3.43 and 3.44 where

$$y_i = 1n(I_{Di}) \quad (3.94)$$

and

$$x_i = \left(\frac{V_{Fi} - RS \cdot I_{Di}}{V_t}\right) \quad (3.95)$$

for i = 1 to 10. The value of RS used in Equation 3.95 is that found from the results of Method 1 given in Equation 3.91. For a linear regression data fit using this transformed data, the matrix equation below is generated from Equation 3.47

$$\begin{bmatrix} 10 & 390.3876 \\ 390.3876 & 16911.02 \end{bmatrix} \cdot \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} 2.420368 \\ 528.9597 \end{bmatrix}$$

(3.96)

This equation is solved for b and m where

$$b = -9.909661 \quad (3.97)$$

and

$$m = 0.2600451 \quad (3.98)$$

From Equations 3.48 and 3.49, values for the parameters IS and N are calculated, and presented with RS where

$$IS = \exp(b) = 49.69228\ \mu A \quad (3.99)$$

$$N = \frac{1}{m} = 3.84554 \quad (3.100)$$

and

$$RS = 40.20590\ m\Omega \quad (3.101)$$

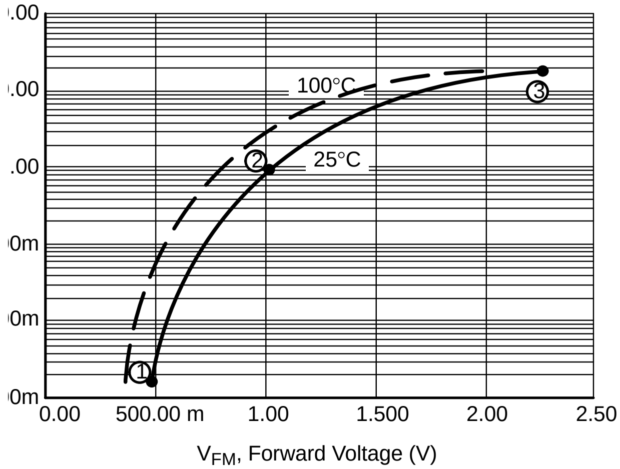


Figure 64. MURH840CT Forward–bias DC Data for the Three–point I–V Method

Table 7. Data for the Three–Point I–V Method

Point	Diode Forward Current (I_D) (A)	Diode Forward Voltage (V_F) (V)
1	1.0 m	0.423
2	1.0	0.973
3	30.0	2.390

3. *Method 3 (Nonlinear curve fit with fixed IS)*. The maximum instantaneous reverse current (I_R) at 25°C is specified on the data sheet as 10 μA. Since this current is modeled by IS, it is necessary to keep its value less than 10 μA in order to satisfy the maximum IR specification. Thus, for this method, IS is set to 3.0 pA which is close to the value at 25°C and 400 volts shown in the plot of reverse current versus reverse voltage of Figure 69. From Table 8, the 10 I–V data points are transformed into y data points with Equation 3.52, and the two current functions with Equations 3.53 and 3.54 where

$$Y_i = V_{Fi} \quad (3.102)$$

$$f_1(I_{Di}) = V_t \cdot 1n\left(\frac{I_{Di}}{IS} + 1\right) \quad (3.103)$$

and

$$f_2(I_{Di}) = I_{Di} \quad (3.104)$$

for i = to 10. The value of IS used in equation (103) is 3.0 *A*. For a nonlinear curve fit using this transformed data, the matrix equation below is generated from Equation 3.57

$$\begin{bmatrix} 1.202756 & 43.00399 \\ 43.00399 & 2276.01 \end{bmatrix} \cdot \begin{bmatrix} a_2 \\ a_2 \end{bmatrix} = \begin{bmatrix} 5.401513 \\ 224.1998 \end{bmatrix}$$

(3.105)

This equation is solved for a_1 and a_2 where

$$a_1 = 2.986477 \tag{3.106}$$

and

$$a_2 = 42.0777m \tag{3.107}$$

From Equations 3.58 and 3.59, values for the parameters N and RS are calculated, and presented with IS where

$$N = a_1 = 2.986477 \tag{3.108}$$

$$RS = a_2 = 42.07777\ m\Omega \tag{3.109}$$

and

$$IS = 3.0\ \mu A \tag{3.110}$$

Table 8. Forward Current Versus Forward Voltage for the MURH840CT

Diode Forward Current (I_D) (A)	Diode Forward Voltage (V_F) (V)
1.0 m	0.423
10.0 m	0.512
100.0 m	0.663
1.0	0.973
5.0	1.408
10.0	1.706
15.0	1.925
20.0	2.104
25.0	2.256
30.0	2.390

To determine which of these three sets of extracted parameters best simulates this example device, the diode forward voltage is calculated from the SPICE model and compared to the data sheet value at the same current. These calculations are performed for each set of the extracted parameters and listed in Table 9 where V_{FC} is the voltage calculated from

$$V_{FC} = N \cdot V_T \cdot 1n\left(\frac{I_D}{IS} + 1\right) + RS \cdot I_D \tag{3.111}$$

and ε is the error in percent between the data sheet voltage V_F and V_{FC} calculated as

$$\varepsilon = \frac{V_{FC} - V_F}{V_F} \cdot 100\% \tag{3.112}$$

At the bottom of this table are the error functions (E_2) calculated from the errors generated by the results of each method in accordance with Equation 3.35. From an examination of these values, it is clear that the parameters extracted with Method 1 produce the lowest value for E_2. However, the forward voltage errors for currents beyond 1.0 A (where the device is most likely to be used) are significantly larger than those produced by the parameters extracted with the other two methods. The parameters extracted by Method 2 yield lower errors in this range even though it has a larger error function. However, the value of IS extracted by this method is too large to satisfy the data sheet specification for maximum I_R. The parameters extracted by Method 3 produce the largest error function of the three while having errors above 1.0 A comparable to those of Method 2. The value of IS used in this method does, however, meet the maximum I_R specification. Plots of the dc characteristics resulting from each of the three methods are shown in Figures 65(a) through 65(c). Both data sheet and calculated voltages are plotted versus current to illustrate the closeness of data fit.

At this point in the extraction process, a decision must be made as to which of the three parameter sets should be used to most accurately model the device in a practical and realistic manner. If the maximum I_R specification (at both 25°C and 150°C) can be waived, then the parameters extracted with Method 2 should be used. If this waiver cannot be exercised, which is often the case, then the parameters from Method 3 should be used in spite of the inaccuracy of the data fit for currents below 1.0 A. For this example, the maximum I_R specification will be observed and the parameters extracted by Method 3 will be used.

Table 9. SPICE Diode Voltages and Percent Error for All Three Methods

MURH840CT Data Sheet Values		Method 1		Method 2		Method 3	
		IS = 3.255398 μA RS = 40.20590 m N = 2.854546		IS = 49.69228 μA RS = 40.20590 m N = 3.845540		IS = 3.0 μA RS = 42.07777 m N = 2.986477	
I_D (A)	V_F (V)	V_{FC} (V)	ε (%)	V_{FC} (V)	ε (%)	V_{FC} (V)	ε (%)
1.0 m	0.423	0.4230000	0.0000	0.3033381	–28.2889	0.4488392	6.1086
10.0 m	0.512	0.5930908	15.8380	0.5283141	3.1864	0.6368089	22.4236
100.0 m	0.663	0.7666325	15.6308	0.7604325	14.6957	0.8083743	21.9267
1.0	0.973	0.9727604	–0.0246	1.0255167	5.3974	1.0240414	5.2458
5.0	1.408	1.2523701	–11.0533	1.3463608	–4.3778	1.3166286	–6.4894
10.0	1.706	1.5045580	–11.8078	1.6163086	–5.2574	1.5805403	–7.3540
15.0	1.925	1.7355133	–9.8435	1.8576529	–3.4986	1.8222380	–5.3383
20.0	2.104	1.9577755	–6.9498	2.0872862	–0.7944	2.0548409	–2.3365
25.0	2.256	2.1752744	–3.5783	2.3105025	2.4159	2.2824602	1–1729
30.0	2.390	2.3897603	–0.0100	2.5296600	5.8435	2.5069275	4.8924
–		E_2 = 0.0914765		E_2 = 0.1155167		E_2 = 0.1203895	

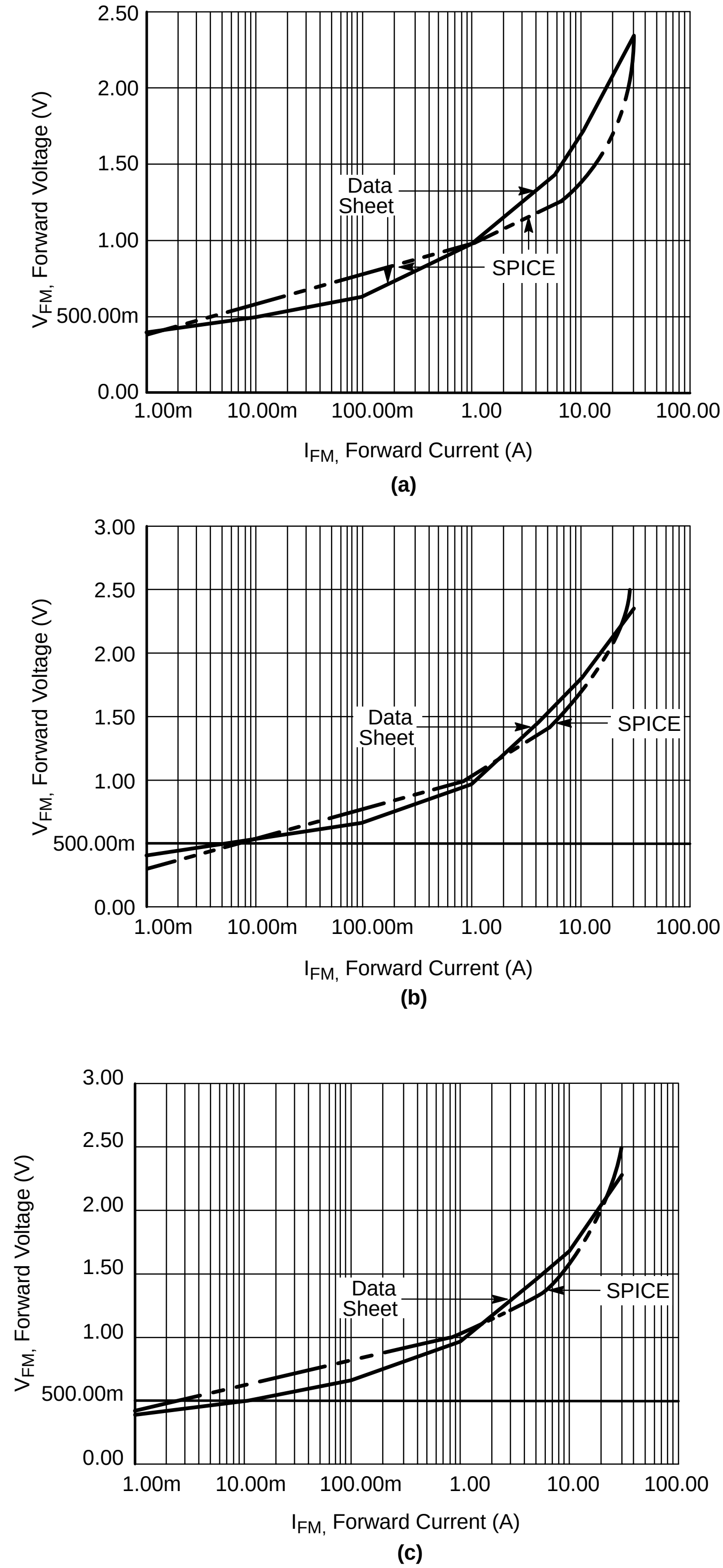


Figure 65. (a) MURH840CT DC Characteristics from Method 1;
(b) MURH840CT DC Characteristics from Method 2;
(c) MURH840CT DC Characteristics from Method 3

Junction Capacitance Characteristics (CJO, VJ, M, FC)

The reverse–bias C–V curve for the MURH840CT is shown in Figure 66. Data from this curve will be used in each of the two methods for extracting the parameters CJO, VJ, M and FC.

1. *Method 1 (Three–point C–V method)*. From Figure 66, a point is selected at the lower voltage end of the C–V curve as point 1. Points 2 and 3 are selected at the upper voltage end as indicated. The capacitance and reverse voltage values corresponding to these points are shown in Table 10.

   From Equations 3.60 through 3.64, the following junction capacitance parameters are computed from this data:

$$M = 0.4012080 \qquad (3.113)$$

$$VJ = 0.2159243\text{ V} \qquad (3.114)$$

$$CJO = 152.7494\text{ pF} \qquad (3.115)$$

and

$$FC = 0.5 \qquad (3.116)$$

Table 10. Data for the Three–Point C–V Method

Point	Diode Reverse Voltage (V_R) (V)	Diode Junction Capacitance (C_j) (pF)
1	0.01	150.0
2	70.0	15.0
3	100.0	13.0

2. *Method 2 (Linear regression with fixed VJ)*. From Figure 66, the C–V data presented in Table 11 is obtained. These 11 C–V data points are transformed into *x–y* data points with Equations 3.68 and 3.69 where

$$y_i = 1n(C_{ji}) \qquad (3.117)$$

and

$$x_i = 1n\left(1 + \frac{V_{Ri}}{VJ}\right) \qquad (3.118)$$

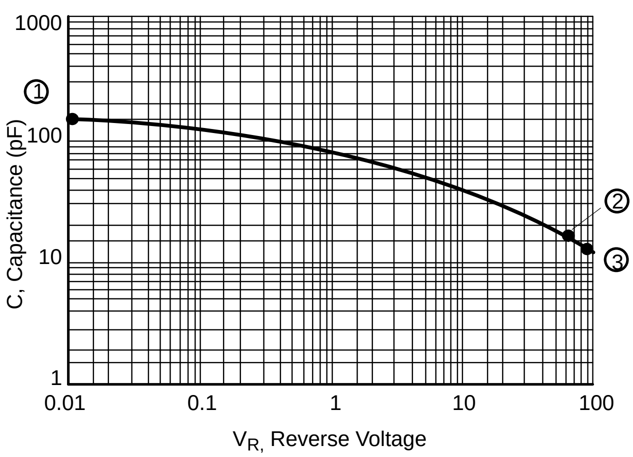


Figure 66. MURH840CT Reverse–bias Junction Capacitance Data for the Three–point C–V Method

for i = 1 to 11. The value of VJ used in Equation 3.118 is that found from the results of Method 1 given in Equation 3.114. For a liner recession data fit using this transformed data, the matrix equation below is generated from Equation 3.47:

$$\begin{bmatrix} 11 & 31.17795 \\ 31.17795 & 133.9153 \end{bmatrix} \cdot \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} -261.1229 \\ -758.7108 \end{bmatrix} \qquad (3.119)$$

Table 11. Junction Capacitance Versus

Reverse Voltage for the MURH840CT	
Diode Reverse Voltage (V_R) (V)	Diode Junction Capacitance (C_j) (pF)
0.01	150.0
0.10	130.0
0.40	100.0
0.70	90.0
1.00	80.0
2.50	60.0
5.50	40.0
10.00	30.0
30.00	20.0
70.00	15.0
100.00	13.0

This equation is solved for b and m where

$$b = -22.58128 \qquad (3.120)$$

and

$$m = -0.4082637 \qquad (3.121)$$

From Equations 3.72 and 3.73, values for the parameters CJO and M are calculated, and presented with VJ and EC where

$$CJO = \exp(b) = 155.9821\text{ pF} \qquad (3.122)$$

$$M = -m = 0.4082637 \qquad (3.123)$$

$$VJ = 0.2159243\text{V} \qquad (3.124)$$

and

$$FC = 0.5 \qquad (3.125)$$

To determine which of these two sets of extracted parameters best simulates this example device, the diode junction capacitance is calculated from the SPICE model and compared to the data sheet value at the same voltage. These calculations are performed for each set of parameters and listed in Table 12 where C_{jC} is the calculated capacitance

$$C_{jC} = CJO \cdot \left(1 + \frac{V_R}{VJ}\right)^{-M} \qquad (3.126)$$

and ε is the error in percent between the data sheet capacitance C_j and C_{jC} calculated as

$$\varepsilon = \frac{C_{jC} - C_j}{C_j} \cdot 100\% \qquad (3.127)$$

At the bottom of this table are the error functions (E_2) calculated from the errors generated by the results of each method. From an examination of these values, it is clear that the parameters extracted with Method 2 produce the most accurate data fit to the actual device. Therefore, for this example, the parameters extracted by Method 2 will be used. Plots of the junction capacitance characteristics resulting from both methods are shown in Figures 67(a) and 67(b). Both data sheet and calculated capacitance are plotted versus reverse voltage to illustrate the closeness of data fit.

Table 12. SPICE Diode Junction Capacitances and Percent Error for Both Methods

MURH840CT Data Sheet Values		**Method 1**		**Method 2**	
		CJO = 152.7494 pF VJ = 0.2159243 V M = 0.4012080 FC = 0.5		CJO = 155.9821 pF VJ = 0.2159243 V M = 0.4082637 FC = 0.5	
V_R (V)	C_j (F)	C_{jc} (F)	ε (%)	C_{jc} (F)	ε (%)
10.0 m	150.0 p	150.0000 p	0.0000	153.1256 p	2.0837
100.0 m	130.0 p	131.1195 p	0.8611	133.5354 p	2.7195
400.0 m	100.0 p	100.3089 p	0.3088	101.6770 p	1.6769
700.0 m	90.0 p	85.5457 p	–4.9492	86.4700 p	–3.9222
1.0	80.0 p	76.3538 p	–4.5577	77.0247 p	–3.7191
2.5	60.0 p	55.3101 p	–7.8164	55.4806 p	–7.5323
5.5	40.0 p	41.0343 p	2.5857	40.9452 p	2.3631
10.0	30.0 p	32.5062 p	8.3539	32.3030 p	7.6767
30.0	20.0 p	21.0384 p	5.1922	20.7476 p	3.7379
70.0	15.0 p	15.0000 p	0.0000	14.7049 p	–1.9674
100.0	13.0 p	13.0048 p	0.0370	127170 p	–2.1769
E_2 = 0.021063, E_2 = 0.018759					

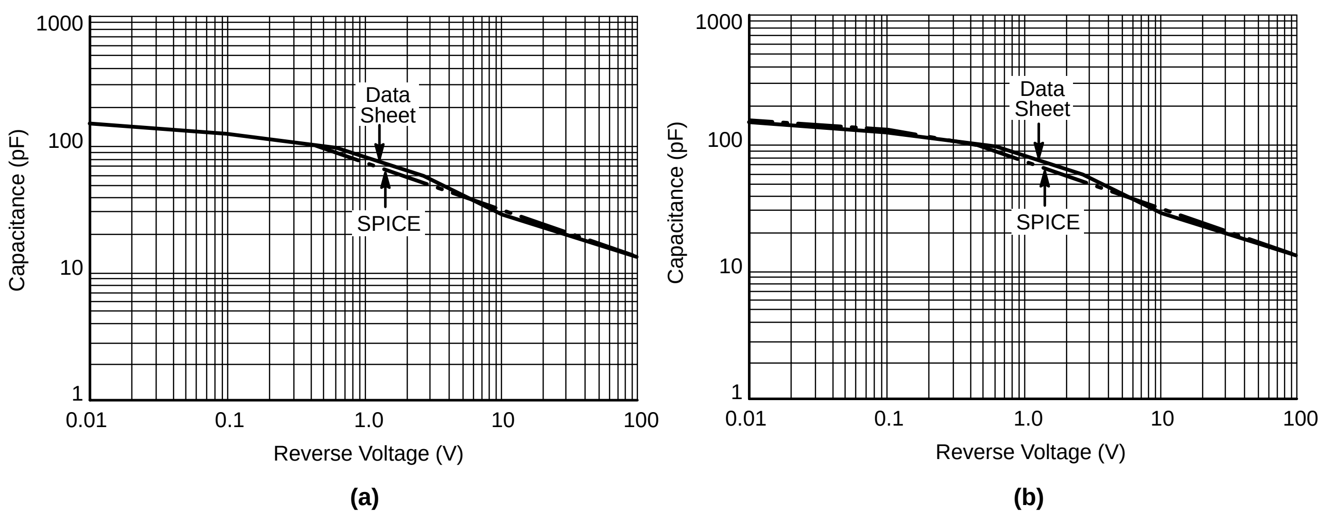


Figure 67. (a) MURH840CT Junction Capacitance Characteristics from Method 1; (b) MURH840CT Junction Capacitance Characteristics from Method 2

Reverse Recovery Time Characteristics (TT)

The typical value for the reverse recovery time t_{rr} for the MURH840CT is 24.6 ns. This value is specified at a forward–bias current (I_F) of 1.0 A and a current slew–rate (di/dt) of 50 A/s. The function equation for the forward transit time parameter TT generated from Equation 3.79 is shown below

$$f(TT) = 15.129 \cdot 10^{-9} - TT \cdot \exp\left(-\sqrt{\frac{20 \cdot 10^{-9}}{TT}}\right) = 0 \quad (3.128)$$

From Newton's method, the value of TT that satisfies this equation is found to be

$$TT = 32.96688 \text{ nseconds} \quad (3.129)$$

Temperature Characteristics (EG, XTI)

The temperature characteristics of the MURH840CT are illustrated in Figures 68 and 69. Even though only three temperature points are given on these two plots, there is sufficient data available to allow the parameters EG and XTI to be extracted with Methods 2 and 3. Since information about the small–scale voltage temperature coefficient is not known, Method 1 will not be used.

1. *Method 2 (Large–scale TCV method)*. The large–scale TCV is obtained from the forward–bias I–V curve of Figure 68. At a forward current (I_D) of 1.0 A, the diode voltages at 25°C (T_o) and 100°C (T_1) are read as 0.973 V ($V_F(T_o)$) and 0.776V ($V_F(T_1)$), respectively. The value of the large–scale TCV is calculated from Equation 3.84 where

$$TCV = -2.62667 \text{ mV/°C} \quad (3.130)$$

Selecting the value for the energy gap parameter EG as 1.1.1 eV which is typical for silicon, the value for the parameter XTI is computed from equation (85) as

$$XTI = 23.702105 \quad (3.131)$$

2. *Method 3 (Reverse leakage current method)*. In Figure 69, the reverse leakage current (I_R) versus reverse voltage at three temperatures is plotted. At a reverse voltage of 400 V, the reverse currents at 25°C (T_o) and 150°C (T_1) are about 3.0 μA ($I_R(T_o)$) and 300.0 μA ($I_R(T_1)$), respectively. Again for EG of 1.11 eV, the value for the parameter XTI is computed from Equation 3.87 as

$$XTI = 3.710244 \quad (3.132)$$

In Table 13, the values of EG and XTI from each method are listed along with the diode voltage calculated at a current of 1.0 A for 27°C and 100°C, and the *saturation* current parameter IS calculated at 150°C. With the value of XTI obtained from Method 2, it is seen that the value of the TCV is close to that computed from the data sheet given in Equation 3.130. However, IS and, consequently, I_R calculated at 150°C is found to be much larger than the maximum value of I_R specified at this temperature which is 500 μA. By using I_R data over temperature to extract XTI as was done with Method 3, IS at 150°C is found to be very close to I_R at the same temperature as shown in Figure 69. The value of the TCV computed with this XTI is. however, not as accurate that produced by the XTI of Method 2. Again, the specification for maximum IR will be observed and XTI computed by Method 3 given in Equation 3.132 will be used.

Table 13. Comparison of Temperature Characteristics

Parameter or Variable	Method		
	2	3	Units
EG	1.11	1.11	eV
XTI	23.702105	3.710244	
V_F (27°C)	1.024041	1.024041	V
V_F (150°C)	0.827068	0.966944	V
TCV	–2.698264	–0.782152	mV/°C
IS (150°C)	2.989445 m	300.0	A

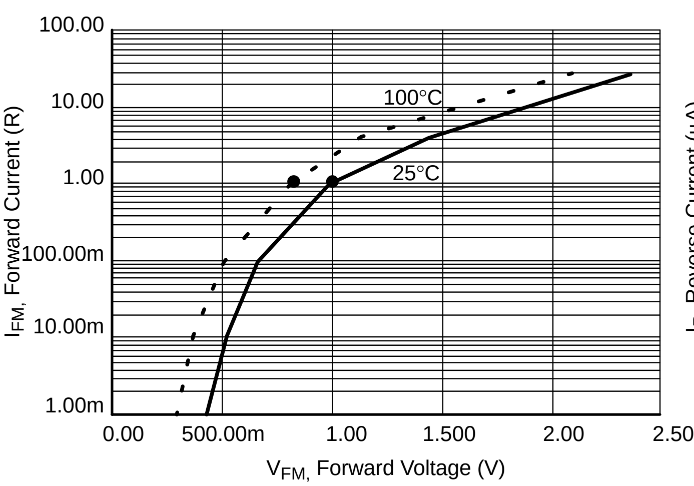


Figure 68. MURH840CT Forward–bias Voltage Temperature Characteristics

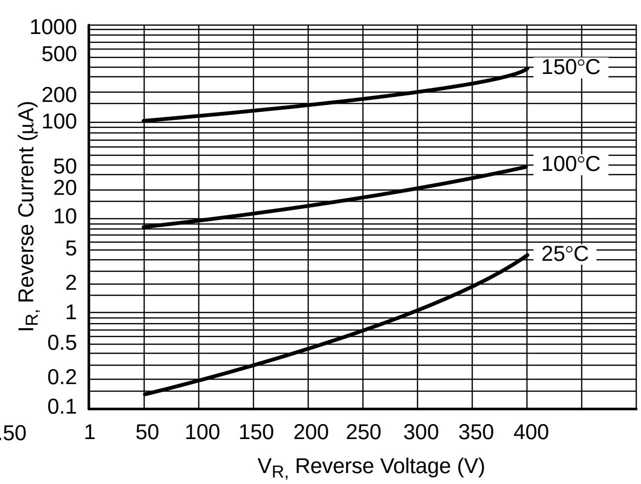


Figure 69. MURH840CT Reverse Leakage Current Temperature Characteristics.

Reverse Breakdown Characteristics (BV, IBV)

The maximum dc blocking voltage V_R(max) for this device is read from the data sheet as 400.0 V. Using a guard–band multiplier k_{BR} of 1.3, the breakdown voltage parameter BV is calculated from Equation 3.88 to be

$$BV = 520.0\ V \qquad (3.133)$$

For a maximum simulation temperature of 150°C, the current at the breakdown voltage IBV is calculated from Equation 3.89 as

$$IBV = 4.279703\ A \qquad (3.134)$$

Summary

To summarize the extraction process, all SPICE parameters for the MURH840CT are listed in Table 14. Characteristic plots generated by the SPICE diode model using these parameters are shown in Figures 70a) through (d).

The SPICE diode model has been shown to consist of a set of equations and parameters derived from the theory developed for an ideal diode. Certain modifications of these equations allow the model added accuracy in the simulation of real diodes. Although the model has some limitations, it can be used to realistically model the behavior of most semiconductor pn junction devices which include low~current integrated circuit diodes, high–current rectifiers, varactor diodes, Zener diodes, and Schottky barrier diodes.

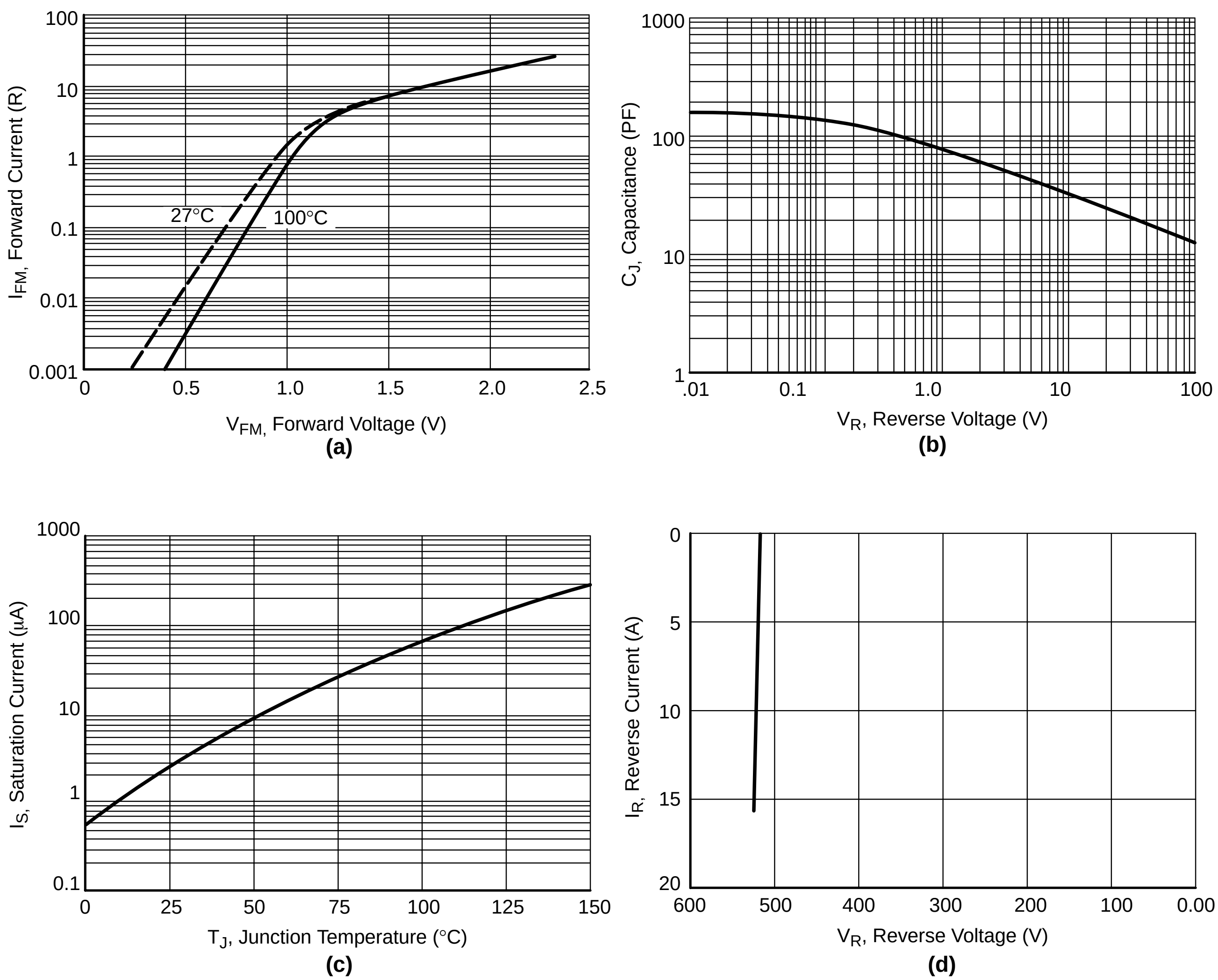


Figure 70. (a) MURH840CT SPICE Model Forward Voltage DC Characteristics at 27°C and 100°C; (b) MURH840CT SPICE Model Reverse–bias C–V Characteristics; (c) MURH840CT SPICE Model Reverse Leakage Current Characteristics; (d) MURH840CT SPICE Model Reverse Breakdown Characteristics

Table 14. SPICE Diode Model Parameters for the MURH840CT

Name	Parameter	MURH840CT Value	Units
IS	Saturation current	3.0 µ	A
RS	Ohmic resistance	42.07777 m	ohm
N	Emission coefficient	2.986477	
TT	Forward transit time	32.96688 n	sec
CJO	Zero–bias junction capacitance	155.9821 p	F
VJ	Contact potential	0.2159243	V
M	Junction capacitance grading exponent	0.4082637	
EG	Energy gap	1.11	eV
XTI	IS temperature exponent	3.710244	
KF	Flicker noise coefficient	0	
AF	Flicker noise exponent	1.0	
FC	CJ forward–bias coefficient	0.5	
BV	Reverse breakdown	520.0	V
IBV	Current at BV	4.279703	A

References

1. A. S. Grove, *Physics and Technology of Semiconductor Devices,* John Wiley and Sons, Inc., New York, NY, 1967.
2. A. Bar–Lev, *Semiconductors and Electronic Devices,* Prentice–Hall International, Inc., London England, 1979.
3. U. S. Muller and T. I. Kamins, *Device Electronics for Integrated Circuits*, John Wiley and Sons, Inc., New York, NY, 1977.
4. B. G. Streetman, *Solid State Electronic Devices. Second Edition*, Prentice–Hall Inc., Englewood Cliffs, NJ, 1980.
5. SPICE 2 User s Guide. University of California Berkeley, Electronics Research Labs, Berkeley, CA.
6. P. Antognetti and G. Massobrio, *Semiconductor Device Modeling With SPICE,* McGraw–Hill Book Co., Inc., New York, NY, 1988.
7. C. F. Gerald, *Applied* Numerical *Analysis,* Addison–Wesley Publishing Co.. Reading, MA, 1970.
8. B. Miller, "Curve Fitting Made Easy", *RF Design,* Vol. 13, No. 6, June 1990, pp. 27 to 32.
9. B. A. Leinfelder, "Improving *PN* Junction Reverse Recovery Measurements", *PCIM,* Part 1, January 1988, pp. 41 to 44; Part 2, February 1988, Pp. 43 to *45;* Part 3, March 1988, pp. 52 to *54.*
10. Rectifiers and Zener Diodes Data Book, Technical Information Center, ON Semiconductor, Inc., 1987.

Chapter 4

Rectifier Diode Specifications and Ratings

Rectifier Diode Specifications and Ratings

Understanding the electrical and thermal characteristics discussed earlier is helpful in design work. However, diodes are tested by the manufacturer and ratings are based on compliance to given specifications. Therefore, a thorough understanding of the meaning and significance of specifications and the development of ratings is mandatory if devices are to be used reliably and economically.

Specification formats for rectifier diodes are established by the rectifier JEDEC[1] committee for registered part numbers, that is, 1NXXXX numbers. Although non–registered house–numbered parts do not always conform to the registration specification requirements, the JEDEC specifications generally are used as guidelines. Furthermore, adherence to letter symbols and definitions as developed by JEDEC is widespread, although they have changed in the past and will undoubtedly change in the future. The information presented is based on the present status of the standard ANSI–EIA–282–A–1989[1], which is referred to in this chapter as “The Standard.” The Standard was developed by the JEDEC Committee on Rectifier Diodes and Thyristors and approved by ANSI in September 1989.

The Standard contains an extensive section on testing. Most of the tests described are not used for high volume production, but are chosen to insure that a rectifier diode is capable of meeting its ratings in a low–frequency, usually 60 Hz powerline application. These tests are intended to be used to verify the specifications of JEDEC registered part numbers and may be considered “referee” tests. In this chapter, tests from the Standard are referred to as “Standard Tests.”

Since the Standard has changed from previous issues, past practices are noted because many rectifiers used today were registered years ago. In addition, background information is given so that future changes may be understood.

Table 15 indicates the JEDEC symbols and terms used. The terms are also defined on the waveforms of Figure 71. The waveform sketches are typical of operation in a half–wave rectifier circuit driven from a sine wave source, which is the traditional basis of rating rectifier diodes. Two cycles of normal operation are shown followed by a half–cycle current surge of peak value I_{SFM}. All values of terms are referenced to the zero baseline. Instantaneous values of i_F, v_F and p_F are shown at arbitrary times t_1 and t_2 on the waveforms; the other terms apply at specific points.

In general, the subscripts carry a consistent meaning throughout the waveforms. M stands for maximum or peak, S stands for surge, a nonrepetitive event, F stands for forward, and (AV) stands for an average value, sometimes referred to as a dc value. R has a dual meaning: when following the primary symbol (i.e. I, V, P or T) such as in I_R it stands for reverse; when preceding the subscript M, it stands for repetitive. A few exceptions occur to the nomenclature scheme and some terms can’t be shown on the waveforms; these cases are described in Table 15.

[1]Joint Electron Device Engineering Council of the Electronic Industries Association (ETA) and National Electrical Manufacturers Association (NEMA).

Table 15. JEDEC Letter Symbols for Rectifier Specifications

	Voltage		Current		Power Dissipation		Junction Temperature
	Forward	Reverse	Forward	Reverse	Forward	Reverse	
Total RMS	$V_{F(RMS)}$	$V_{F(RMS)}$	$I_{F(RMS)}$	$I_{R(RMS)}$	–	–	–
Alternating Component RMS	V_f	V_r	I_f	I_r	–	–	–
DC (No AC)	V_F	V_R	I_F	I_R	P_F	P_R	T_J
DC (With AC)	$V_{F(AV)}$	$R_{V(AV)}$	$I_{F(AV)}$ I_0(1)	$I_{R(AV)}$	$P_{F(AV)}$	$P_{R(AV)}$	$T_{J(AV)}$
Instantaneous Total Value	V_F	V_R	i_F	i_R	p_F	p_R	t_J
Maximum (Peak)	V_{FM}	V_{RM}	I_{FM}	I_{RM}	P_{FM}	P_{RM}	T_{JM}
Specific Maximum Peak Total[2]							
Repetitive	–	V_{RRM} V_{RWM}	I_{FRM} I_{FSM}	I_{RRM}	–	–	T_{JRM}
Non–repetitive	–	V_{RSM}	$I_{FM(OV)}$	I_{RSM}	–	–	*T_{JSM}
Special cases[3]							T_{JFRM} T_{JRWM} T_{JFSM}

1. Rectified Output Current, Average 180° Conduction Angle. 50 or 60Hz, Sine Wave Input
2. Primarily used for ratings, see text
3. Primarily used in developing current ratings, see Figure 71 and text.

Historically the major use of rectifiers is in the conversion of 60 Hz supply line voltage to direct current; consequently, the earliest JEDEC specifications consisted simply of a current rating, a voltage rating, and some measurement of reverse current and forward voltage. Presently thermal and reverse recovery data is also required on all rectifier diodes because of their widespread use in power electronics. Graphs of the characteristics specified, showing the effect of voltage, current, and temperature, are frequently given on manufacturers' data sheets.

Various specifications (generally referred to as specs) in common use are examined in this section. Their main purpose is to insure that the rectifier meets its current and voltage ratings and to insure interchangeability of parts from different manufacturers.

Temperature

Temperature is the fundamental limiting factor of rectifier diode power handling capability. Manufacturers establish current and voltage ratings to insure long life using various criteria, but temperature is implicit in the ratings.

The maximum and minimum storage temperatures (T_{stg}) simply define the temperature range allowable during storage. Excessively hot temperatures can damage the die–to–case bond or damage the package or alter electrical or thermal characteristics, which result in a field failure when the diode is placed into operation. Excessively cold temperature can cause cracking of the chip with subsequent voltage degradation or catastrophic failure. Minimum operating temperatures are usually the same as minimum storage temperatures.

The maximum repetitive peak temperature T_{JRRM} is the temperature used to establish V_{RRM}. It usually is stated on data sheets as $T_{J(MAX)}$.

The maximum peak values T_{JSM} and T_{JFRM}, though generally not stated, are factors in determining the surge ratings and the average current ratings with capacitive or high form factor loads which cause high peak power dissipation. The peak temperature limits are useful in calculating operating limits under unusual pulse or transient conditions; manufacturers usually provide these limits upon request. Temperature ratings during forward current conduction are usually higher than those applicable when a reverse voltage is present, because the combined stress of temperature and voltage is the most sensitive chip failure mode.

Thermal Resistance

Older rectifier diodes do not have thermal resistance specified since the current rating alone as a function of diode reference temperature is a sufficient guide to operation in 60 Hz rectifier currents. Diodes in power electronics circuits, however, experience a variety of waveforms not covered by commonly published current derating curves. In order for designers to compute average junction temperature, thermal resistance is required and has been a JEDEC requirement for a number of years.

Peak junction temperature, particularly in applications where the diode handles short duration pulses, can be considerably higher than the average temperature. To calculate peak temperature, the superposition technique is commonly used with a curve of transient thermal impedance or thermal response as described in Chapter 2. Not all classes of rectifier diodes are presently required by JEDEC to specify transient thermal impedance, but manufacturers usually publish a thermal response curve on data sheets.

Some data sheets contain a specification called "apparent thermal resistance" which applies for a particular periodic waveform of forward current. Apparent thermal resistance is calculated such that the peak temperature is obtained by multiplying the apparent thermal resistance by the average power produced by the waveform.

A typical thermal resistance or transient thermal impedance test fixture uses the test diode's forward voltage at a low "metering" current as an indicator of junction temperature. A switching circuit alternately applies a high current pulse followed by the metering current. For obtaining thermal resistance, the duty cycle of the high current pulse is nearly 100%. A measurement of the diode "metering voltage" permits thermal resistance to be calculated. Two approaches find use in determining thermal response. The first uses high–power single–shot pulses of varying width followed by a measurement of the metering voltage. The second applies a pulse long enough to achieve steady state conditions; after the current drops to the metering level, the cooling curve is recorded by capturing the change in forward voltage vs. time. Thermal characteristics tests require judgment regarding the most appropriate technique and circuit to use as well as attention to numerous details, which are discussed in "The Standard."

Voltage Ratings

Voltage ratings are not a characteristic; they cannot be measured but are assigned by the device manufacturer and apply over the entire temperature operating range of the device. Ratings should not be exceeded under any circumstances. All 'repetitive ratings are referenced to a half–wave 60 Hz rectifier circuit using a resistive load.

The first three ratings to be discussed are generally specified with the same numbers for a particular rectifier. Although the theoretical basis of peak reverse voltage ratings is the maximum reverse power generation permissible before thermal runaway[2] is reached, these ratings are frequently limited for other reasons, such as an abrupt change in slope of the V/I curve, surface effects, or instability of the V/I curve. Voltage ratings are generally well below avalanche breakdown voltage, but they may be a limiting factor at low temperatures.

[2]Voltage ratings generally apply when maximum rated forward current is flowing. Under this condition the thermal resistance, junction–to–ambient is necessarily rather low which prevents thermal runaway. However, thermal runaway can occur well below rated voltage if $R_{\theta JA}$ and T_A are relatively high.

The equipment designer has the responsibility for determining that none of the reverse voltage ratings of the device are exceeded under any possible combination of operating or environmental conditions. If voltages exceeding any of the reverse voltage ratings are applied to the rectifier diode, a markedly increased probability of failure exists. The ratings do not imply any safety factor. Figure 71 illustrates the various voltages applicable to rectifier diodes.

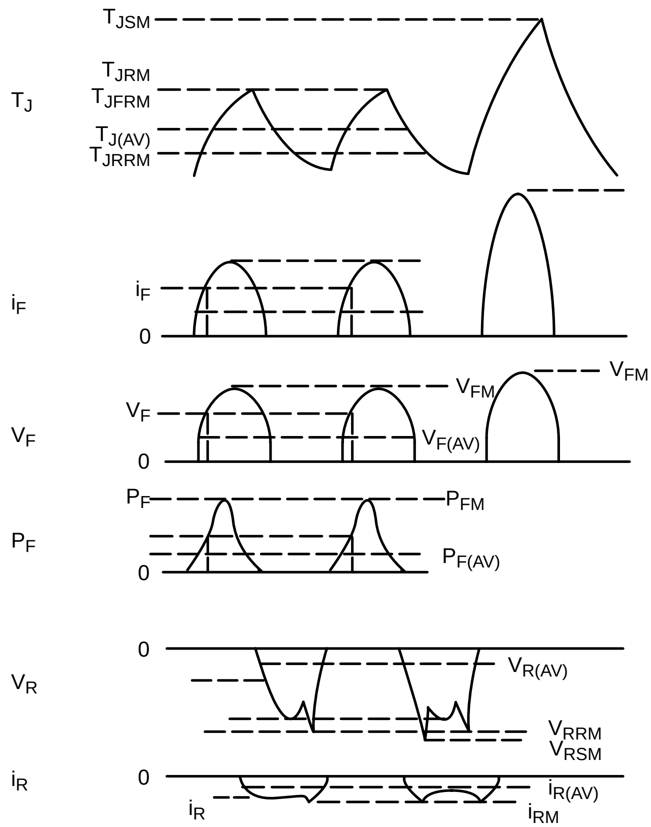


Figure 71. Illustration of JEDEC Letter Symbols. Sketches show Reverse and Forward Voltage and Current and Junction Temperature Excursion Resulting from the Developed Power

Peak Repetitive Reverse Voltage (V_{RRM})

The rated peak repetitive reverse voltage (V_{RRM}) is the maximum allowable instantaneous value of the reverse voltage, including all repetitive transient voltages, but excluding all nonrepetitive transient voltages that occurs across a rectifier diode. To verify the rating, a pulse of 100 ms at a 60 Hz rate is used. The peak repetitive reverse voltage which occurs in a circuit is caused by rectifier diode properties in conjunction with circuit constants and is, to some extent, under the control of the equipment designer. V_{RRM} is a periodic voltage which includes commutation spikes and inductive kicks which occur each cycle.

Peak Working Reverse Voltage (V_{RWM})

The rated peak working reverse voltage (V_{RWM}) is the maximum allowable instantaneous value of the reverse voltage that occurs across a rectifier diode, excluding all repetitive and nonrepetitive transient voltages. The input voltage to the circuit must be such that the peak inverse voltage applied to the rectifier does not exceed rated V_{RWM}. The rating is generally based upon operation in a half–wave 60 Hz rectifier circuit with resistive load and applies over the entire operating temperature range.

DC Reverse Blocking Voltage (V_R)

The rated maximum dc reverse blocking voltage is the maximum allowable de reverse voltage, excluding all repetitive and nonrepetitive transient voltages, across a rectifier diode. The rating applies over the entire operating temperature range specified. Values for V_R are usually the same as V_{RWM}. but V_R may be restricted to operation at a lower temperature than V_{RWM}.

Peak Nonrepetitive Reverse Voltage (V_{RSM})

The rated peak nonrepetitive reverse voltage (V_{RSM}) is the maximum allowable instantaneous value of reverse voltage across the rectifier, including all nonrepetitive transient voltages but excluding all repetitive transient voltages. The rating applies over the entire operating temperature range. The nonrepetitive peak reverse voltage occurs as a random circuit transient, which may originate within the equipment or be externally generated. Transient voltages may be minimized by transient voltage surge–suppression components, as discussed in Chapter 9.

Reverse Current

Reverse leakage current is so low with the majority of silicon p–n junction rectifier diodes that it plays no role in design. In these cases, reverse leakage specifications are used to verify voltage ratings. However, the processing used to reduce lifetime in the manufacturing of ultrafast diodes causes a marked increase in reverse leakage current. Schottky diodes also naturally exhibit high reverse leakage. Consequently, designs using fast diodes must include a thermal runaway analysis and accurate reverse current data is essential. The following reverse current specifications are found on manufacturer's data sheets.

Maximum or Peak Reverse Current (I_{RRM})

The maximum reverse current specified is the peak reverse current through the rectifier with a half cycle of 60 Hz sinusoidal rated reverse voltage (V_{RWM}) applied at maximum operating temperature – i.e., the case or lead temperature at which the forward current is derated to zero, If reverse power is negligible, the maximum operating temperature is equivalent to T_{JRRM}.

In production, I_{RRM} is generally measured at room temperature using a pulse technique because it can be done quickly even though high–temperature leakage is not predicted well by low–temperature measurements. The manufacturer relies primarily on test data correlating high– and low–temperature leakage and on sample testing of production lots to insure compliance to the registered high temperature limits.

Average Reverse Current $I_{R(AV)}$

The average reverse current specified is the value of reverse leakage current averaged over one complete 60 Hz cycle under the stated conditions of reverse voltage and case temperature. Rated forward current may also be applied as an option. This specification serves as a comparison of the rectifier to an ideal switch of zero leakage in the open position.

Measurement of $I_{R(AV)}$ does not blend with a production environment and is not specified on modem diodes. Older registered parts often specified $I_{R(AV)}$ usually with forward current applied. Its measurement requires use of a dynamic load test circuit which simulates operation in a half–wave rectifier circuit at rated conditions, If accurate dynamic load test data is to be retrieved, the diode must be mounted on a substantial heat sink having some external means of controlling the case temperature to within a few degrees. If temperature control is not used, the inevitable differences in forward power dissipation between parts will cause significant changes in case and junction temperature as different parts are tested. The resulting data would be useless because of the strong dependence of I_R upon T_J. Consequently I_{RRM} is the preferred reverse current specification.

DC Reverse Current (I_R)

The specified dc reverse current is the value of reverse current through the rectifier under stated conditions of rated dc reverse voltage (V_R) and case temperature. At one time, the temperature specified was required by JEDEC to be at the maximum operating temperature, but, because of thermal runaway problems, it may now be specified at a lower temperature. Sometimes dc reverse current at room temperature is added to diode specifications by manufacturers, but it is measured in production using a pulse.

Reverse Breakdown Voltage V($_{BR}$)

Breakdown voltage is not specified on rectifier diodes unless the diode is a controlled avalanche type. For the nonavalanche diode, verification of reverse voltage ratings is done by means of a reverse leakage current test

Controlled avalanche diodes have two values of “breakdown” voltage specified.

The first is a minimum breakdown specified at a low current chosen to place operation in the breakdown region near the “knee.” It may be performed using either a dc, ac, or pulsed signal source, but production testing favors use of a short pulse at a low duty cycle. Normal circuit operating voltages should be below the minimum breakdown voltage.

The second breakdown test is designed to verify the peak reverse power dissipation specified for a controlled avalanche rectifier. The standard test consists of superimposing a transient voltage to a diode which is biased to rated V_{RWM}. The series impedance and transient source voltage are adjusted until the product of reverse voltage and reverse current equals the specified peak reverse power dissipation of the diode. In production, a constant current pulse is applied to all diodes even though the higher voltage units will be stressed beyond their published ratings because adjustments are too labor intensive and fraught with chances for error.

Forward Voltage

In order to establish current ratings as a function of reference temperature, a curve of forward voltage (V_F) vs. forward current is required. To control V_F a specification at one point is required for JEDEC registered part numbers. Manufacturers generally follow the JEDEC requirements for house–numbered parts and often provide supplemental information.

Maximum (Peak) Forward Voltage (V_{FM})

The specified maximum forward voltage is the peak forward voltage across the rectifier under the same stated conditions of current and case or lead temperature used for the forward current rating (I_O). V_{FM} may be measured using either a 60 Hz half sine wave continuous current or a pulse technique to keep junction heating negligible. Using the half sine wave, the temperature of the heat sink must be controlled by external means so that variations in power dissipation between parts do not influence case temperature, which is to be held to that specified for I_O.

With the pulse method, no heat sink or complicated temperature control is necessary and, consequently, it is more practical to use in production testing. A pulse width under 1 ms at a duty cycle under 2% is specified. Testing at the peak current for which the rectifier is specified, i.e., πI_O is required to insure that the power dissipation will always be below a certain limit, in order that temperature limits (T_{JFRM} or T_{JRRM}) will not be exceeded when operating at rated conditions.

The pulse test is performed at 25°C (room temperature). The limit assigned to V_{FM} when measured with the pulse test is different than that assigned with the 60 Hz test at elevated case temperature in order to account for the temperature coefficient of forward voltage, θ_V. Since parts from a particular product line exhibit a consistent θ_V, a room temperature V_{FM} test assures high temperature behavior.

Average Forward Voltage ($V_{F(AV)}$)

The specified average forward voltage is the forward voltage across the rectifier averaged over one complete cycle under the same stated conditions of average forward current and temperature used for establishing the current rating of the part. $V_{F(AV)}$ is measured in a circuit similar to that used for V_{FM} except that an average reading dc meter is used instead of a peak reading meter.

As previously discussed, temperature control is a problem with the 60 Hz load test; furthermore, this measurement does not necessarily guarantee that the power dissipation will be below the limit necessary to keep temperature limits from being exceeded. This problem is illustrated by the sketch of Figure 72. Two rectifiers having different forward characteristics are shown by curves 1 and 2. Both have the same $V_{F(AV)}$ but the rectifier of curve 2 has a higher V_{FM}; it will have higher power dissipation than rectifier 1. As a result, $V_{F(AV)}$ is no longer a required JEDEC spec and is seen only on older data sheets.

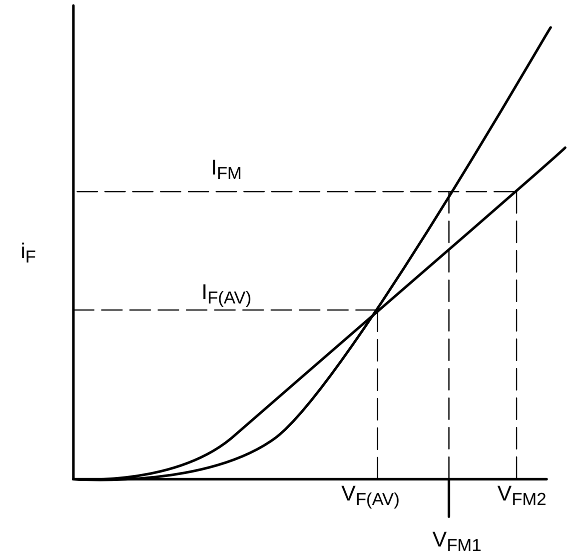


Figure 72. Two Rectifier Forward Voltage Curves Showing Difficulty of Controlling Rectifier Power Dissipation by Limiting Average Voltage

DC Forward Voltage (V_F)

The dc forward voltage is the forward voltage across the rectifier for a given dc current and case temperature. This spec appears on older data sheets, but it does not adequately control V_{FM} and requires case temperature control for accuracy.

Forward Power Dissipation

For convenience of the designer, data sheets usually show a plot of average forward power dissipation versus average forward current for various conduction angles. Data can be obtained by measurement or values may be obtained by using one of several methods of calculation. The methods trade off accuracy for simplicity and range from the watt per amp "rule of thumb"[3] to integration techniques requiring the use of a computer.

Power dissipation may be measured by utilizing the thermal resistance of the heat sink system. A thermocouple is attached to a convenient place on the rectifier case or the heat sink, and the ac current is applied to the input of a suitable rectifier circuit, preferably the one intended for end use. The temperature and average current levels are recorded after thermal stability is achieved. Next, dc current is applied and its level adjusted until the temperature equals that obtained during the previous measurement. The forward drop is now easily measured and used to compute power. This measurement technique is particularly useful when operation is desired at frequencies where the transient power losses of the diode must be considered.

[3]The watt per amp "rule of thumb" is based on the assumption that for any given value of average current, the average forward voltage is one volt. This approximation is conservative at small forward currents and optimistic at the high values. it does, however, provide a reasonable guess of average power without knowing the forward voltage characteristic, providing the conduction angle is close to 1800.

Calculation of ac power losses is complicated by the nonlinear V–I characteristic of semiconductor diodes. If diodes were to have a constant voltage drop with current, the power waveform would naturally be a sine wave for sine wave currents. At low currents, the diode drop varies little with current; consequently, the power is nearly sinusoidal. At high currents, the resistance of the diode becomes dominant, producing a power waveform which approaches the sine–squared waveshape, typical of a resistance. To show this behavior, Figure 73 compares actual diode power waveforms at a low and a high level to sine and sine squared functions.

The most accurate way to solve for average power dissipation is to use a computer in a piecewise integration. The computer is programmed to break the current sine wave into a series of small equal increments. Power for each increment is found by programming the computer to multiply V_F – i_F data obtained by interpolation of data points. The average power dissipation is the sum of these increments divided by the number of increments times the period of the waveform. Accuracy improves with the number of increments used. As a guide, average power dissipation will be accurate to better than 1% if the wave is broken into 20 equal increments and *25* data points are used to define the forward voltage curve.

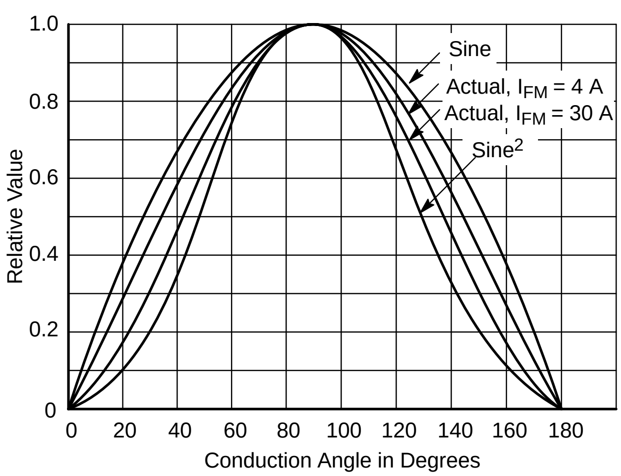


Figure 73. Actual Diode Power Waveforms Compared to Sine Squared Waveforms for Half–Sine Wave Currents

Another method of obtaining average power dissipation is to approximate the forward voltage curve with an offset voltage and a straight line as shown in Figure 74; thus, $V = V_O + mi$, where m is the dynamic resistance of the diode.

Since

$$i = I_M \sin \varpi t$$

$$P_{F(AV)} = \frac{1}{2\pi}\int_O^{\pi} iv d\theta$$

$$= \frac{1}{2\pi}\left[\int_O^{\pi} V_O I_M \sin\theta + \int_O^{\pi} m I_M^2 \sin^2\theta\right] d\theta \quad (4.1)$$

Integration and substitution yield

$$P_{F(AV)} = \frac{V_O I_M}{\pi} + \frac{m I_M^2}{4} \quad (4.2)$$

The offset voltage–straight line approximation is fairly accurate when the peak forward voltage is in the linear region of the V_F – i_F curve. In the nonlinear or exponential region, this approach can be used if a different slope and offset are determined for each value of forward current. Such a determination is an ideal task for a computer. The result is nearly as accurate as piecewise integration, but the program requires considerably less computer time than piecewise integration.

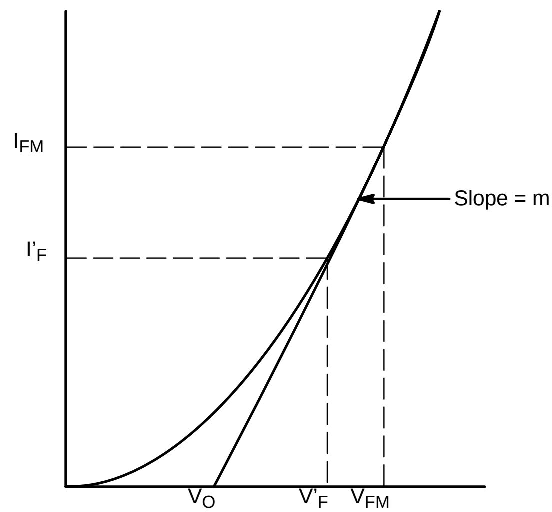


Figure 74. Approximation of a Diode Forward Characteristic for Accurate Calculation of Power Loss when Handling a Sinusoidal Current Waveform

The computer is fed V_F – i_F data and programmed to linearly interpolate between points as in the piecewise method. Based on the sine wave peak value and a percentage of peak value, the computer solves for two V_F values, i.e., for I_{FM} and I'_F. V_{FM} and V'_F are determined from the diode forward voltage curve as shown in Figure 74. Using these two values of forward voltage the slope and offset voltage are found from:

$$m = \frac{V_{FM} - V'_F}{I_{FM} - I'_F} \quad (4.3)$$

and

$$V_O = V_{FM} - I_{FM} m \quad (4.4)$$

Values calculated from Equations 4.3 and 4.4 are used in Equation 4.2 to find power at each current level. Best accuracy generally occurs when $I'_F = 0.9\ I_{FM}$ since the major portion of the sine wave's energy is then between I'_F and I_{FM}.

Of the methods used to calculate the power dissipation in a rectifier, the easiest and most conservative method makes the assumption that the power dissipation waveform is sinusoidal. Thus, the average power dissipation is found as follows:

$$P_{F(AV)} = \frac{1}{2\pi}\int_{O}^{\pi} P_M \sin\theta d\theta = \frac{P_M}{\pi} \qquad (4.5)$$

where $P_M = I_{FM} \times V_{FM}$.

For convenience in calculating power dissipation, Figures 75 and 76 permit form factor and peak to average values to be found for a number of commonly used waveforms.

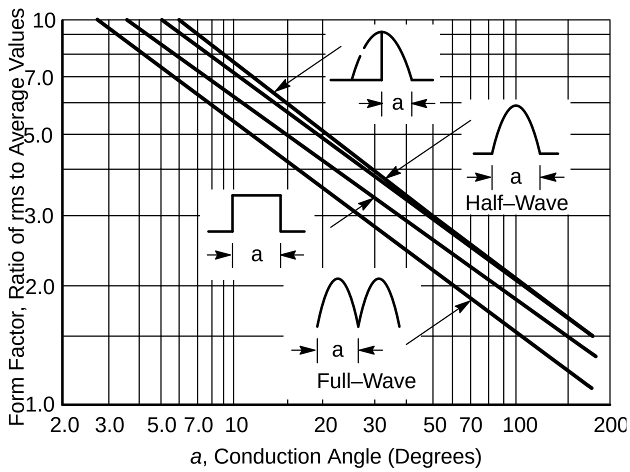


Figure 75. Form Factor for Various Current Waveforms

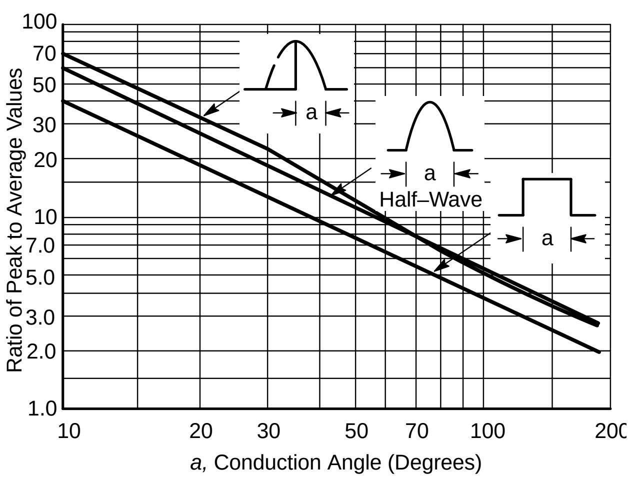


Figure 76. Peak to Average Ratios for Various Current Waveforms

Current Ratings

Of paramount importance are a rectifier diode's current ratings. Current ratings are conditional ratings – dependent upon the diode's case or lead temperature and the waveform being handled.

Several temperature limits play roles in limiting the current that a rectifier diode may handle. Some are indicated on Figure 71, which shows the junction temperature of a rectifier diode operating in a resistively loaded, half–wave, single–phase circuit under steady state conditions followed by a current surge. In addition, electro–migration may set a limit upon the current which a diode can conduct on a continuous basis.

The repetitive peak junction temperature T_{JFRM} sets one limit upon rectifier current ratings. If exceeded, the die bond may be weakened, resulting in an eventual field failure. The limit of T_{JFRM} depends upon the manufacturers' assembly technique. The temperature caused by the circuit depends upon the amplitude and waveshape of the pulse train and the transient thermal impedance of the rectifier.

Another limit is set by T_{JRRM} (often called $T_{J(max)}$), which is the temperature used for determining the voltage rating. Note that T_{JRRM} is lower than $T_{J(AV)}$ for sine wave rectification shown in Figure 71. In practice the values of T_{JRRM} and $T_{J(AV)}$ are close together.

A third limit on the rectifier diode's current rating is the temperature at which the lead material from the die to the terminal fuses or becomes excessively hot. Lead temperature is a function of the lead material resistivity and the rms value of forward current.

The junction temperatures, T_{JFRM} and T_{JRRM}, the lead fusing temperature and electro–migration must all be considered in determining a current rating. Since one or the other will limit current at different operating conditions, manufacturers often provide curves to aid in determining satisfactory operating limits.

A final temperature limit is the peak surge junction temperature, T_{JSM} which is somewhat above the other temperature limits. The surge is allowed only a few times (few being defined by JEDEC as less than 100) during a rectifier diode's life. The surge temperature is the basis for the I_{FSM} rating.

Average Forward Current (I_O)

It is customary, as well as a prior JEDEC requirement, to rate rectifiers in terms of average current delivered to a resistive load at a specified case temperature in a 60 Hz half–wave circuit, unless another frequency is specified. (The 1989 version of Standard 282 indicates that the ratings are to apply for input frequencies from 50 to 400 Hz). The rated current is defined as I_O. The usual basis for this rating is that t_J below T_{JRRM} when V_{RWM} is applied. The reference temperature used for the I_O rating is presently required by JEDEC to be chosen approximately 50°C plus or minus 25°C lower than T_{JRRM}.

It is safe to assume that I_O is also an rms limit when estimating current ratings for capacitive loads or other loads having a high form factor (ratio of tins to average current); however, such an assumption may unduly restrict the usefulness of a particular rectifier. Form factor and peak–to–average ratio data are shown in Figures 75 and 76 as functions of conduction angles for commonly encountered rectifier waveforms. The data are useful in establishing current ratings for conditions other than the ones specified.

Figure 77 is an example of current derating data for a stud– or case–mounted rectifier. The dc (rms) limit is called out at 39.3 A. A check of the curve data against the data of Figures 75 and 76 reveals that the flat portion of the curves shows the same rms limit; that is, for $I_{FM}/I_{F(AV)} = 20$, read a = 28° from Figure 76. At this angle, read F ≈ 3.9 from Figure 75; therefore, $I_{F(AV)} = 39.3/3.9 \approx 10$ A, which agrees with curve 6 on Figure 77. Consequently, derating curves for waveforms other than the ones shown could easily be determined by using the data of Figures 75 and 76 to obtain the rms limit and by using the given derating curves as a guide.

Sine wave derating data is based upon t_J = rated T_{JRRM} a short time after cessation of forward conduction. The dc derating curve, curve 1 on Figure 77 and the square wave derating curve, curve 2, are based upon T_{JRFM} equaling rated T_{JRRM} because the full reverse voltage may appear immediately after forward conduction ceases. Note that the square wave curve, curve 2, is very close to the sine wave resistive load derating data; however, the average current limit is at 28 A rather than 25 A, since $I_{(RMS)}/I_{(AV)} = 1.41$ for a square wave.

Derating data for a lead–mounted rectifier is shown in Figure 78. Since no rms limit is indicated, it can be assumed that it is greater than (3.9)(4.3) or 18.5 A by using data for $I_{FM}/I_{(AV)} = 20$ from Figures 75, 76 and 78. Therefore, other waveforms having a rms value below 18.5 A could be safely handled by the rectifier. For these other waveforms, it is appropriate to use the techniques outlined in Chapter 2 to check for T_{JFRM}. It should be held to a reasonable limit.

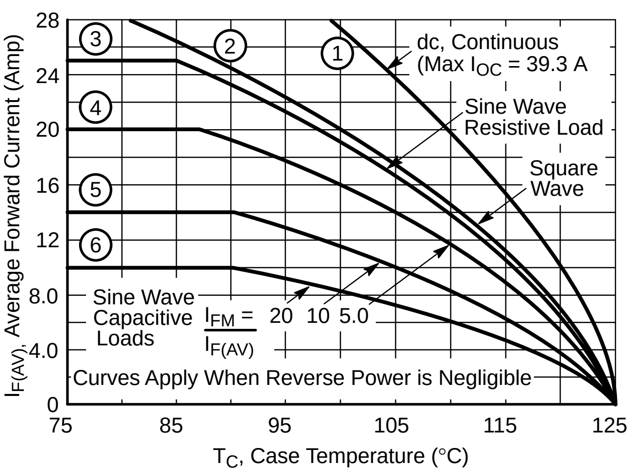


Figure 77. Example of Current Derating Data where RMS Limits are Evident as Flat Portions of the Curves

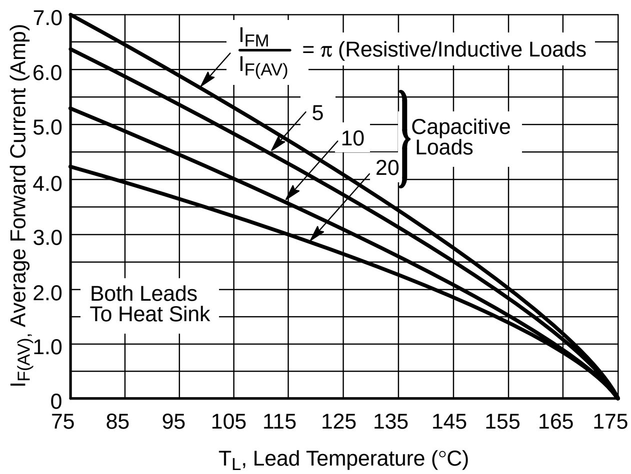


Figure 78. Example of Current Data for a Rectifier which does not show an RMS Limit

Peak Surge Forward Current (I_{FSM})

The maximum allowable surge current (I_{FSM}) is the peak current that the rectifier can safely handle for a minimum of 100 times in its lifetime. It is a nonrepetitive rating in the sense that the surge may not be repeated until thermal equilibrium conditions have been restored.

An illustration of the JEDEC specified conditions is shown in Figure 79. The rectifier is subjected to rated voltage and current until thermal equilibrium is established at the rated case temperature. One half–cycle of surge current, I_{FSM}. is applied, followed by one half–cycle of nonrepetitive rated–voltage, V_{RSM}. If the junction temperature is driven too high by the current, thermal runaway will occur and the rectifier will be destroyed.

In some applications, such as motor control, several cycles of surge current called overload current $I_{(ov)}$ must be handled. To illustrate overload capability, data similar to that of Figure 80 are given on many data sheets. During each cycle of overload, peak reverse voltage must not exceed V_{RWM}. The intent of the JEDEC specification is to have t_J at its rated repetitive limit prior to the nonrepetitive surge. To provide additional information, surge and overload data are often given for the situation when the diode is in a circuit that is turned on at room temperature (the 25°C curves in Figure 80).

Multicycle surge data is used most often to select properly rated circuit breakers and/or to provide sufficient series impedance to prevent diode damage should a load fault occur. Another common use for overload data is to check rectifier operation in a line operated circuit with a capacitive input filter, as shown by the circuit of Figure 81(a). Worst–case current surges occur when the switch is closed as the input sine wave approaches its positive peak (Figure 81(b)). The waveform should be compared to the overload rating curve. In many cases a visual comparison of the current surges is all that is necessary. For example, the rectifier having the surge ratings of Figure 80 will handle two surges of 80 A each when T_J = 175°C or 120 A when T_J = 25°C. The circuit current peaks of Figure 81(b) are 110 A for the first pulse and 60 A for the second one; the other pulses are so low by comparison that they need not be considered. Unless the rectifier is used near its upper temperature limit of 175°C, operation is safe and no further checking is necessary.

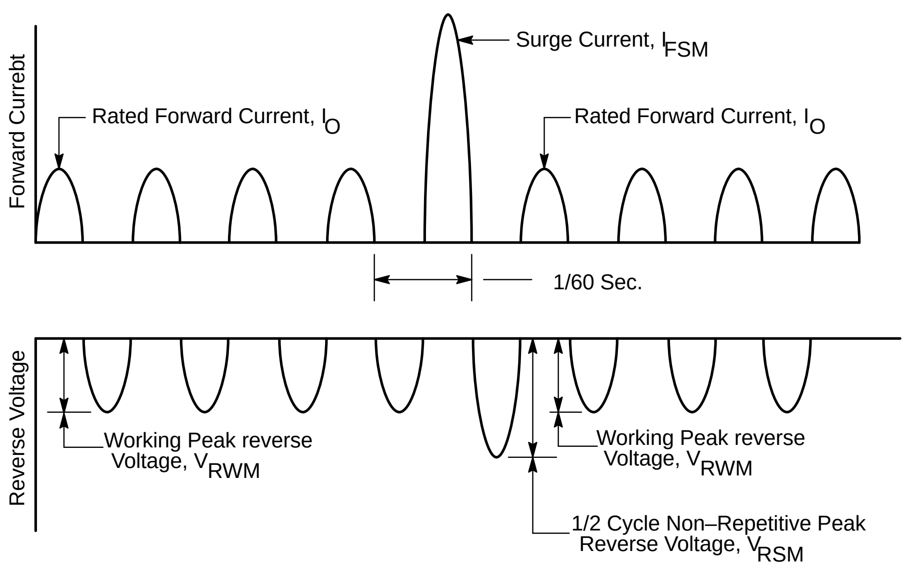


Figure 79. Conditions for the JEDEC Surge Current Specification

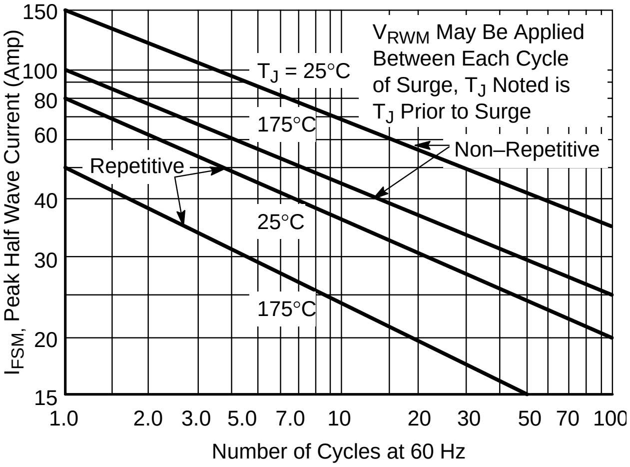
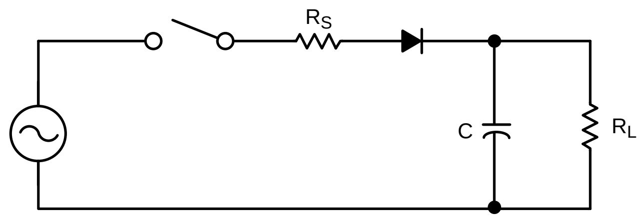


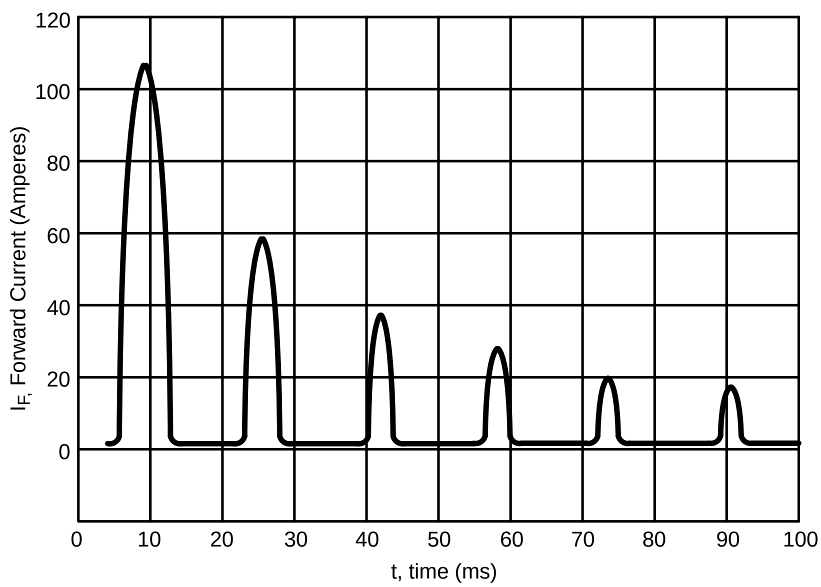
Figure 80. Example of Surge Current Data (applies for MR850 Series)

When more precise information is needed, the simplest way to match the rectifier diode capability to the circuit's demands is to convert the overload surge data of Figure 80 and the waveform produced by the circuit (Figure 81(b)) to equivalent rectangular waves using the thermal modeling techniques discussed in Chapter 2. A satisfactory model is obtained if the amplitude is chosen to equal 70%[4] of the peak value of the actual pulse and the width adjusted to contain the same area. Therefore, the data of Figure 80 can be interpreted as applying a rectangular pulse or a train of pulses having an amplitude of 70% of the sine wave peak amplitude and a width of 5.83 ins.

[4] 70% is chosen because the power pulse under surge conditions will be close to a $\sin^2$ function (see Chapter 2).



(a)



(b)

Figure 81. Start up Transients in a Half–Wave Line Operated Rectifier Circuit: (a) Half Wave Circuit with Large Input Capacitor; (b) Worst–case Turn–on Waveform ($I_{F(AV)}$=100 mA)

Using the methods of calculating instantaneous junction temperature presented in Chapter 2, the maximum temperatures at the end of the power pulse and at the time when the reverse voltage reaches its peak may be calculated for the rectifier data and for the circuit. If the temperature allowed during surge is higher than that produced by the circuit, satisfactory operation can be assumed.

Rectifier diode data sheets sometimes show a subcycle surge rating curve, so called because the time duration is less than the 8.33 ms time base of a 60 Hz wave. The rating is applicable for events which occur while the diode is operating at rated load current and working reverse voltage. However, the diode is not required to block immediately after the subcycle surge current ceases. The rating is useful when selecting a fast acting, current–limiting fuse for protecting the diode during a load fault. Since fuses are rated in terms of I^2t (current squared seconds) an I^2t rating for the diode is often given in place of a subcycle rating curve. Use of subcycle surge data is shown in Chapter 9.

Switching

Originally switching specifications were not required on JEDEC registered rectifier diodes, since the principal applications were at 60 Hz. Because of the widespread use of rectifier diodes in power electronic equipment, reverse recovery characteristics are presently required in order to better assure interchangeability in high–frequency applications. The terms used in switching specifications are defined in Chapter 1.

Reverse Recovery

The test circuit used to obtain data and specification limits for reverse recovery usually conforms to one devised by JEDEC about 1970. As shown on Figure 82, the circuit consists of a means of discharging the energy stored in a capacitor though a series inductor and the diode under test. By properly choosing circuit constants, a half–sine wave pulse of forward current is passed through the diode. The circuit can be modified to also store the charge involved during reverse recovery on a capacitor whose voltage will yield the reverse recovery charge, Q_{RR}.

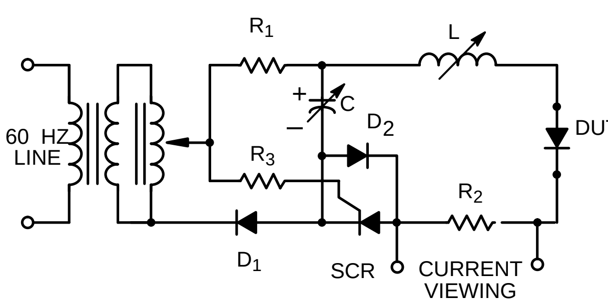


Figure 82. Standard JEDEC Reverse Recovery Test Circuit

The registration requirement states that t_a, t_b, $I_{RM(REC)}$, and Q_{RR} and are all to have maximum limits. The test conditions are also specified. Suggested conditions are $T_C = 25°C$, pulse repetition rate = 60 pulses per second and di/dt = 25 A/µs. The peak forward current must be at least three times I_O; the pulse width is required to be over five times the diode's reverse recovery time to ensure that dud: is fairly linear and nearly constant through zero crossing from $I_{FM}/4$ to $I_{RM(REC)}$ and also to allow enough time for the steady state charge gradient to be established in the diode. The pulse duty cycle is low to ensure negligible heating of the diode.

The circuit of Figure 82 is difficult to adapt to the necessary conditions required to measure ultrafast diodes, i.e., those with t_{RR} under 100 ns. A fundamental limitation is caused by the inductor used to generate the forward current pulse. When the diode enters the high–impedance phase (t_b) during turn–off, the circuit will ring at the resonant frequency determined by the inductor and the diode capacitance[2]. Therefore the circuit cannot measure t_b unless it is longer than the period of the resonant circuit. The latest version of the Standard, therefore, shows an alternate means of driving the diode with square pulses derived from a power MOSFET circuit. The circuit shown in Figure 83, allows a higher di/dt; typical of modern power electronic circuits, to be obtained and circuit inductance can be kept small.

Diode reverse recovery waveforms observed in the circuit of Figure 82 typically appear as shown by one of the sketches in Figure 84. Waveforms obtained in the circuit of Figure 83 are similar, except the forward current pulse is square. At the end of the period denoted as t_a, the diode impedance begins to rise, allowing reverse current to decay. The nature of the decay during the t_b interval depends upon the impurity profile of the diode, the test or circuit electrical conditions and, unfortunately, the loop inductance and parasitic elements of the circuit.

If the waveform decays smoothly in a somewhat exponential manner, as shown by the waveform of Figure 84(a), the diode is said to exhibit "soft recovery." If, at some point during t_b, the waveform exhibits a rapid change in slope, as shown by the other waveforms, the recovery is described as abrupt. Abrupt recovery is undesirable because it often produces circuit ringing, which causes EMI problems in electronic equipment.

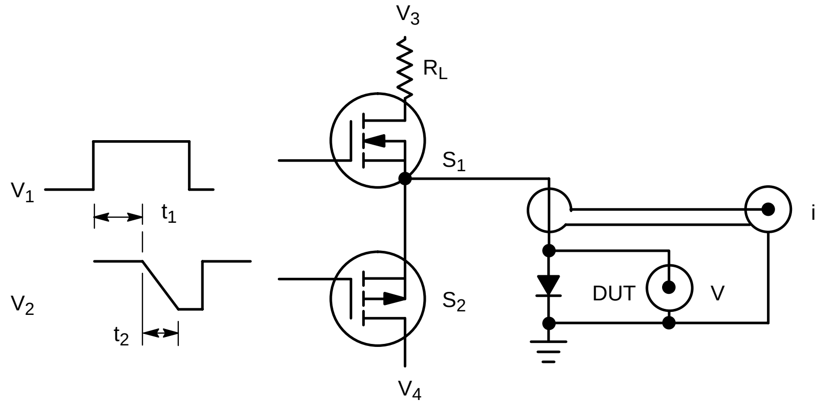


Figure 83. Test Circuit for Ultra–Fast Rectifier Diodes

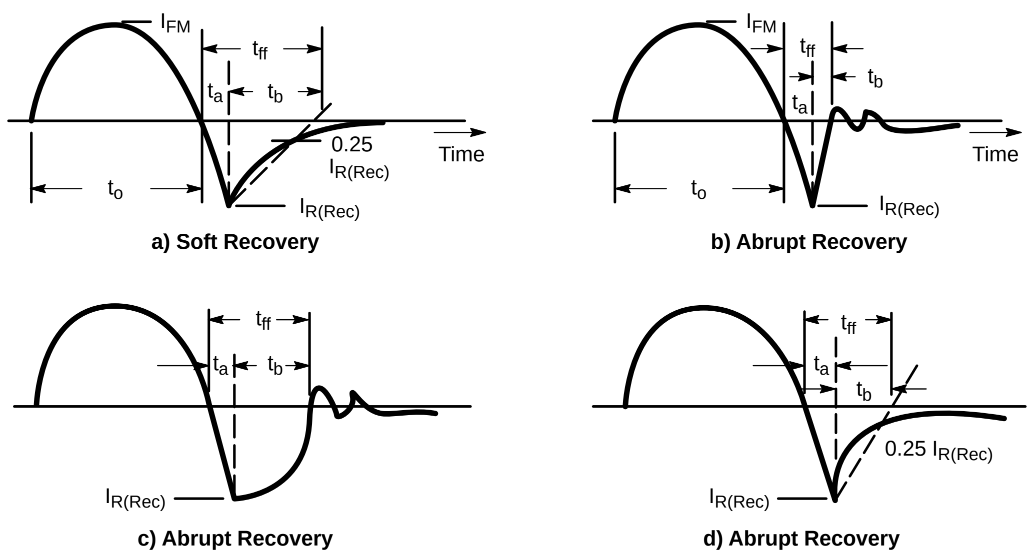


Figure 84. Test Current Waveforms for Various Types of Rectifier Diodes under Test in the Circuit of Figure 82

To aid a designer in judging the nature of the t_b waveform, a recovery softness factor (RSF) is sometimes stated. RSF is the ratio of t_b to t_a; therefore, larger numbers imply softer recovery. Inspection of the waveforms of Figure 84, however, reveals that RSF is of limited value. It does provide a valid comparison of waveforms (a) and (b), but waveforms (c) and (d) yield an RSF comparable to those of the soft waveform of (a), when in fact (c) and (d) are abrupt. If t_b is not accurately measured, a non–trivial task[2], then RSF is of no value at all.

The waveforms of Figure 84 show a long–standing and still–preferred referee method of defining reverse recovery time. The method of measuring t_a is uniform but the measurement of time t_b differs depending upon the character of the waveform. For waveforms which cross the axis because of circuit ringing, such as (b) and (c), t_b is measured from $I_{RM(REC)}$ to the zero crossing. For waveforms approaching the axis asymptotically as shown by (a) and (d), the zero crossing is determined by a straight line drawn from $I_{RM(REC)}$ through the point where i_{RR} has decayed by 75%. The preferred measurement technique must be done manually using an oscilloscope and consequently is not used in a production test environment. It is much more common to see recovery time measurements referenced to 10% of $I_{RM(REC)}$ or to a specified reverse current level.

The difficulty of performing accurate recovery time measurement increases with the speed of the diode. Consequently, a method of measuring the “recovered charge” Q_{RR} without measuring the reverse recovery waveform has been proposed [2] and equipment is available to perform the test electronically. In addition, calculation of recovery time has shown good agreement with carefully measured data for most diodes. The recovered charge is the amount of charge delivered to an external circuit by the diode during turn–off. It is the area under the reverse recovery waveform (see Figure 84). The electronic technique uses a means of transferring the recovered charge to a capacitor where it is stored and whose voltage is easily measured by a voltmeter. The method is described in the Standard and also in MIL–STD–750 C, method 4061.1.

Forward Recovery

Forward switching characteristics are seldom specified on rectifier data sheets even though a diode's turn–on behavior is important in a number of power electronic applications operating at high frequencies. The reason for the absence of forward recovery data is that manufacturers are reluctant to add tests not demanded by design engineers. The importance of having fast turn–on and low dynamic impedance is evidently not fully appreciated.

Although a forward switching test is not required on JEDEC registered diodes, the Standard shows a procedure for measurement and has defined terms. The test circuit is very simple; it is a series loop consisting of a pulsed current source, the diode and a means of measuring current. The current source is usually obtained by adding a series resistance to a voltage generator. The open circuit voltage of the source is required to be 50 V or $3V_{FM}$, whichever is greater. The rise time of the input wave should be fast enough so that it does not affect the diode's performance.

The turn–on voltage wave, with pertinent terms noted, is shown on Figure 85. V_{FRM}, t_{fr}, and the steady state pulse voltage, V_F, are specified. The measurement point V_{FR} for t_{fr} is 1.1 V_F when V_{FRM} between 1.3 V and 10 V When V_{FRM} exceeds 10 V, V_{FR} is specified at $3V_F$. Should V_{FRM} be below 1.3 V, $V_{FR} = V_F + 1/2\ (V_{FRM} – V_F)$.

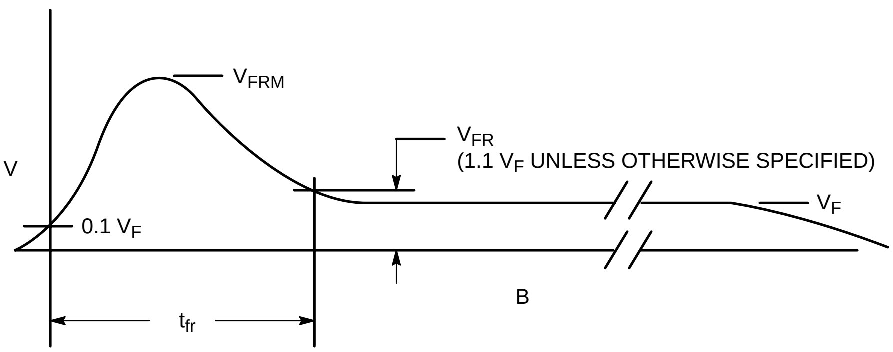


Figure 85. Turn–on Waveform Showing JEDEC Defined Terms.

References

1. "Standard for Silicon Rectifier Diodes," ANSI/EAI–282–A–1989, Electronic Industries Association, 2001 Pennsylvania Aye, N.W. Washington, D.C. 20006
2. "B. Leinfelder, "Improving PN Junction Reverse Recovery Measurements," Pans 1, 2, and 3, "Powerconversion and Intelligent Motion," Jan., Feb., March, 1988.

Chapter 5

Basic Single–Phase Rectifying Circuits

Basic Single-Phase Rectifying Circuits

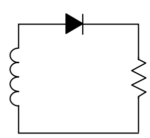
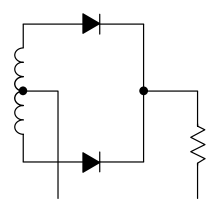
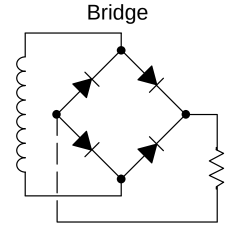
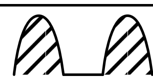
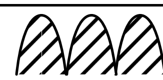
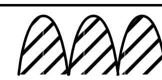
In this chapter, basic characteristics of single–phase circuits with resistive loads are discussed. Polyphase circuits are treated in Chapter 6; filter design follows in Chapter 7. Results of all the calculations in this chapter are presented in Table 16.

Basic Operation

When considering any rectifying circuit, a designer desires to know the magnitude of the direct voltage and current, the regulation of the load voltage, and the efficiency to be expected from the rectifying process. All these values depend upon a number of variables–such as the type of circuit, the constants of the supply, the characteristics of the rectifying unit, and the nature of the load–which complicate the analytical solution. However, certain simplifying assumptions idealize the circuit so that a useful analysis can be made.

The simplest circuit for rectifying single–phase alternating torrent gives half–wave rectification. Such a circuit is indicated in Figure 86(a), where the bold–faced arrow represents the rectifying unit and the direction of conventional current flow. Assume (1) an ideal AC source without resistance, (2) the impressed emf is a pure sine wave, (3) the rectifying unit has zero resistance in the forward direction of current and infinite resistance in the reverse direction, and (4) the load is purely resistive. With these assumptions, let a sine wave alternating voltage as shown in Figure 86(b) be impressed across the input to the rectifying circuit. The output is a rectified half–wave of current as indicated in Figure 86(c), which is of a sine form, since $i = (V_M/R_L) \sin \omega t$. During the second half of the cycle the rectifier blocks current. The current flowing through the load resistance R_L produces an iR_L voltage drop which has a half sine waveform. The voltage waveform across the rectifier is shown by Figure 86(e).

Table 16. Characteristics of Basic Single–Phase Rectifier Circuit with Resistive Loads

	Half–Wave	Full–Wave Center–Tap	Full–Wave Bridge
Rectifier Circuit Connection			
Load Voltage and Current Waveshape Characteristic			
Diode Average Current $I_{F(AV)}/I_{L(DC)}$	1.00	0.50	0.50
Diode Peak Current $I_{FM}/I_{F(AV)}$	3.14	3.14	3.14
Form Factor of Diode $I_{F(RMS)}/I_{(DC)}$	1.57	1.57	1.57
Diode RMS Current $I_{F(RMS)}/I_{L(DC)}$	1.57	0.785	0.785
RMS Input Voltage Per Transformer Leg $V_i/V_{L(DC)}$	2.22	1.11	1.11
Peak Inverse Voltage $V_{RRM}/V_{L(DC)}$	3.14	3.14	1.57
Transformer Primary Rating VA/P_{DC}	3.49	1.23	1.23
Transformer Secondary Rating VA/P_{DC}	3.49	1.75	1.23
Total RMS Ripple, %	121	48.2	48.2
Lowest Ripple Frequency, f_r/f_i	1	2	2
Rectification Ratio (Conversion Efficiency), %	40.6	81.2	81.2

1. $P = I_L^2 R_L \qquad V_L = I_L R_L$

Current Relationships

Since the function of rectification is to convert alternating current to direct current, the equivalent value of the direct current output is of primary interest. Its value is measured by the dc ammeter placed in the circuit of Figure 86(a).

This dc value is the average value of th6 instantaneous rectified current over one cycle or a period corresponding to 2π radians. Average values can be determined by a graphical and arithmetic process by measuring the instantaneous value of current at a series of equally spaced points along the time axis and then dividing the sum of these values by the total number of points in a cycle. A more accurate solution may be obtained by measuring the area under the rectified current pulse and dividing by 2π. The more direct solution is to apply calculus to the problem, thus:

$$I_{DC} = I_{AV} = \frac{1}{2\pi}\int_0^{\pi} I_M \sin \omega t d(\omega t) \qquad (5.1)$$

Solving,

$$I_{DC} = \frac{2I_M}{2\pi} = 0.318\ I_M \qquad (5.2)$$

An ac ammeter inserted in the rectifying circuit of Figure 86(a) gives a different reading from the dc meter first considered. This difference arises because the ac meter registers effective or root–mean–square (rms) values [1] rather than average values. The heating or effective value of a current varies as the square of the instantaneous current. Hence, to determine the effective value of the rectified current graphically, it is first necessary to plot squared values of the instantaneous current as suggested by the dotted i^2 curve on Figure 86(c). The effective area under the dotted i^2 curve may be obtained by the graphical point method, or by calculus:

$$\text{Effective area} = \int_0^{\pi} I_M^2 \sin^2 \omega t d(\omega t)$$

$$= I_M^2 \left[\frac{\omega t}{2} - \frac{\sin 2\omega t}{4}\right]_0^{\pi} = \frac{I_M^2 \pi}{2} \qquad (5.3)$$

and the effective or rms current,

$$\sqrt{\frac{\text{effective area}}{2\pi}} = 0.5\ I_M \qquad (5.4)$$

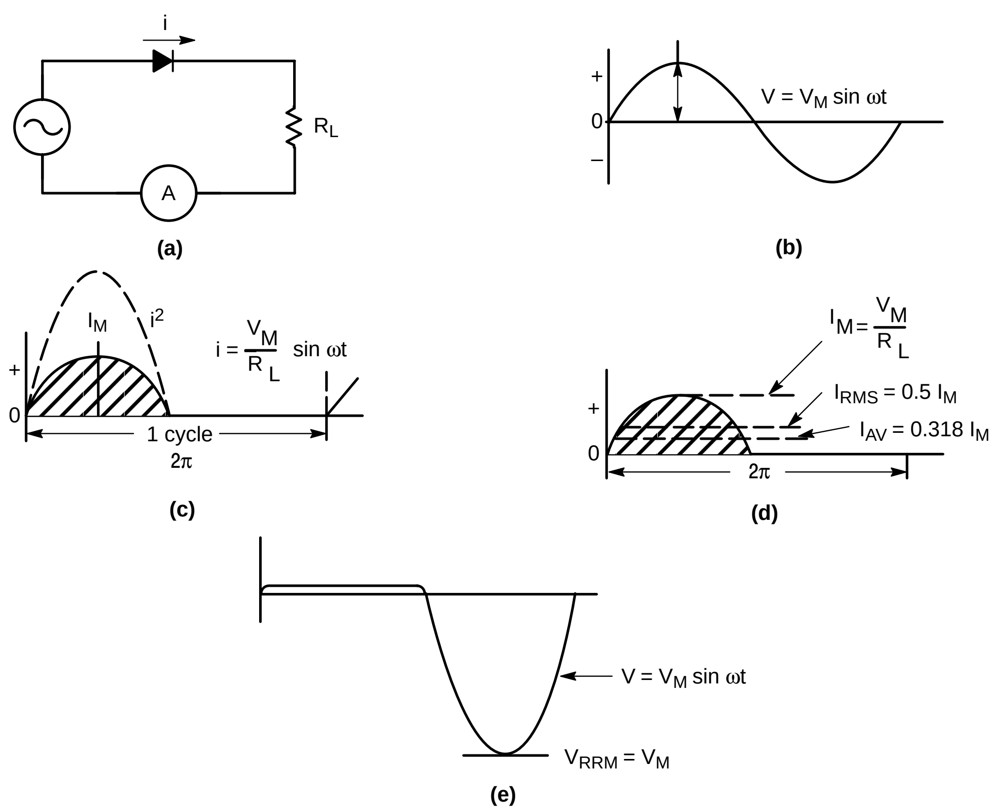


Figure 86. Circuit and Wave Shapes for Half–Wave Rectification: (a) Half–Wave Circuit; (b) Input Waveform; (c) Current Waveform; (d) Current Waveform with Pertinent Points Indicated; (e) Rectifier Voltage Waveform

The average and effective values of rectified current for the half–wave circuit are indicated in Figure 86. The integration of Equations 5.1 and 5.3 follow the standard procedures for obtaining the average and effective, values of periodic functions.:

Half–wave circuits are not used with transformers unless current requirements are small, because the dc component of output current must flow through the transformer. The dc component causes core magnetization and high core losses result. However, the half–wave circuit finds limited use in low–current, direct–line rectification.

Two types of circuits are used for full–wave, single–phase rectification. One circuit uses a transformer with a mid–tap in the secondary winding, as shown in Figure 87(a). The other uses a bridge configuration, shown in Figure 87(b), which requires two extra rectifier diodes, but the secondary requires only half as much winding. Performance is similar except that the bridge diodes are subjected to only half the peak inverse voltage of the center–tap circuit (to be discussed later). Since both half–waves of current pass through the transformer (dividing in the secondary of the center–tap circuit), there is no dc component of flux in the transformer core to increase core losses.

Using the same assumptions made for the preceding half–wave rectifier circuit, the relation between the input and output sides of either full–wave rectifier circuit may be readily calculated. Since both halves of the cycle are rectified, the current and voltage on the input side are normal effective values; rms values on the output side are the same as for a sine wave while the dc or average values are twice that of a half–wave circuit. The relationships are shown in Figure 87(d).

Form Factor

Rectifier circuit efficiencies can be related to a quantity termed form factor (fl. It is the ratio of the heating component of a wave to the dc component:

$$F = \frac{I_{(RMS)}}{I_{(AV)}} \qquad (5.5)$$

Substituting the values from Equations 5.2 and 5.4 into Equation 5.5, the form factor for a half–wave circuit is 1.57. The form factor is the same for each rectifier element in the full–wave systems because each side conducts on opposite half cycles. However, the form factor of the output current of a full–wave circuit is much better. It is found from the values shown on Figure 87; i.e., F = (0.707)1(6.636) = 1.11.

Form factor takes on significance when rectifiers must handle high peak currents at low duty cycles, since the power losses in the diodes and transformers are much higher than encountered with sine wave pulses. Applications where high form factors are the rule occur when capacitive input filters are used, when batteries are being charged, and when rectifiers are used with SCRs in phase control circuitry.

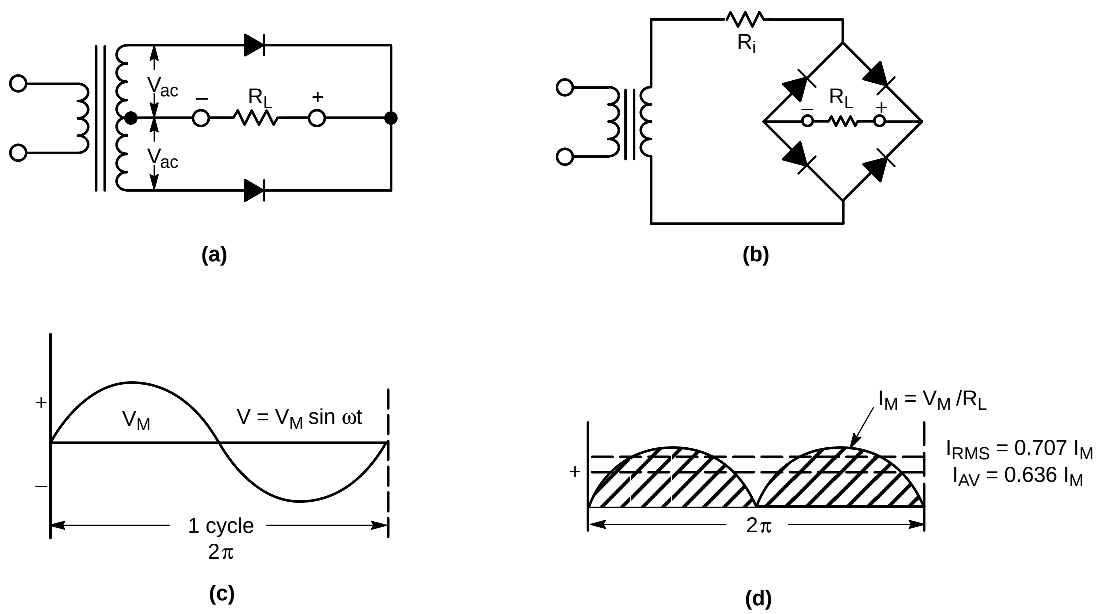


Figure 87. Circuit and Waveshapes for Full–Wave Rectification: (a) Center–Tap Circuit; (b) Bridge Circuit; (c) Input Waveform; (d) Output Waveform

Utilization Factor

Because of the waveforms involved in rectifier circuits, transformers are not used as efficiently as when they handle pure sinusoidal waveforms. A measure of rectifier circuit merit is the utilization factor (UP), defined as the ratio of the dc output power to the transformer volt–ampere rating required by the primary and/or the secondary. For single–phase circuits with resistive loads, the UF can be found by using the relationships between rms and average current. As defined,

$$U_F = \frac{I_{AV} V_{AV}}{I_{RMS} V_{RMS}} \qquad (5.6)$$

For the half–wave circuit:

$$U_F = (0.318\ I_M)\ V_{DC} / (0.51\ I_M)(2.22\ V_{DC})$$
$$= 0.286 \qquad (5.7)$$

For a full–wave circuit, transformer utilization is much improved because conduction is continuous. In the center–tap circuit, although both the secondary windings are present, only one is used at a time. The utilization factor for the secondary is found as

$$V_F = \frac{(0.318\ I_M)(2)V_{DC}}{2(0.5\ I_M)(1.11)V_{DC}} = 0.572$$

In the primary, the whole winding is naturally in continuous use; the input power is, assuming a 1:1 winding for simplicity, $(\sqrt{2}/2)I_M V_{i(RMS)}$.

Therefore,

$$V_F = \frac{(0.318\ I_M)(2)V_{DC}}{(\sqrt{2}\,/\,2)I_M(1.11)V_{DC}} = 0.812$$

The bridge rectifier circuit has the same utilization factor as the primary of the full–wave center–tap circuit. The UP of the single–phase bridge is quite high and is only exceeded by certain polyphase circuits.

However, utilization factor does not tell the whole story in the case of a half–wave circuit since only one half of the sine wave of current is passed, and the windings carry a dc component of current which magnetizes the iron core and increases the core losses. As a result, half–wave rectifiers are only practical for use with a transformer when the current requirement is very small.

Harmonics

When the current in a transformer winding does not have a sinusoidal waveform, harmonic currents are present. The harmonic content can be found by a Fourier analysis of the waveform or a spectrum analysis of current in an operating circuit. For half sine waves the percentage of each harmonic with respect to the fundamental is given in Table 17:

Table 17. Harmonic Percentages of a Half Sine Wave

Harmonic	2nd	3rd	4th	5th	6th
%	21.2	0	4.2	0	1.8

Harmonics cause transformer core loss and hysteresis loss to be higher than with single–frequency operation because these lasses increase with frequency. The extra loss caused by the second harmonic in the resistive loaded rectifier circuit is not high compared to other losses in the transformer and is usually neglected. ‘With other loads, losses caused by harmonics can be appreciable and not only complicate transformer design but also cause trouble with the electrical distribution system.

Ripple Factor

The rectified voltage and current output consists of a series of unidirectional waves or ripples. For some applications these variations are not objectionable, but for others they must be smoothed out by filters. For all cases, the relative magnitude of the ripple is important in the comparison of rectifying circuits. The comparison is made in terms of ripple factor. Ripple factor is the ratio of effective value of the alternating components of the rectified voltage or current to the average value. In equation form ripple factor is

$$r_f = \frac{\text{effective rectified ac load component } I'_{rms}}{\text{average load current } (I_{DC})} \qquad (5.7)$$

Percent ripple is a term used interchangeably with ripple factor and is simply the ripple factor expressed in percent (r_f X 100).

The various components of current associated with ripple factor are illustrated in Figure 88. Figure 88(a) shows the single current pulse of half–wave rectification, and Figure 88(b) shows the splitting of the pulse into a dc component and a ripple component. Figures 88(c) and 88(d) give an actual separation of the components, and Figure 88(e) illustrates the components for full–wave rectification.

In accordance with the preceding definition, the ripple factor is the ratio of the effective current represented by I’ of Figure 88(c) to the dc component shown in Figure 88(d). This ratio may be computed by following the preceding form of calculation. Thus, the instantaneous ac ripple component i' may be represented as

$$i' = i - I_{DC}, \qquad (5.8)$$

and the total rms value of the ripple component I'_{rms} is

$$I'_{rms} = \sqrt{\frac{1}{2\pi}\int_0^{2\pi} (i - I_{DC})^2\, d\theta}$$

$$I'_{rms} = \sqrt{\frac{1}{2\pi}\int_0^{2\pi} (i^2 - 2I_{DC} + i^2_{DC})\, d\theta} \qquad (5.9)$$

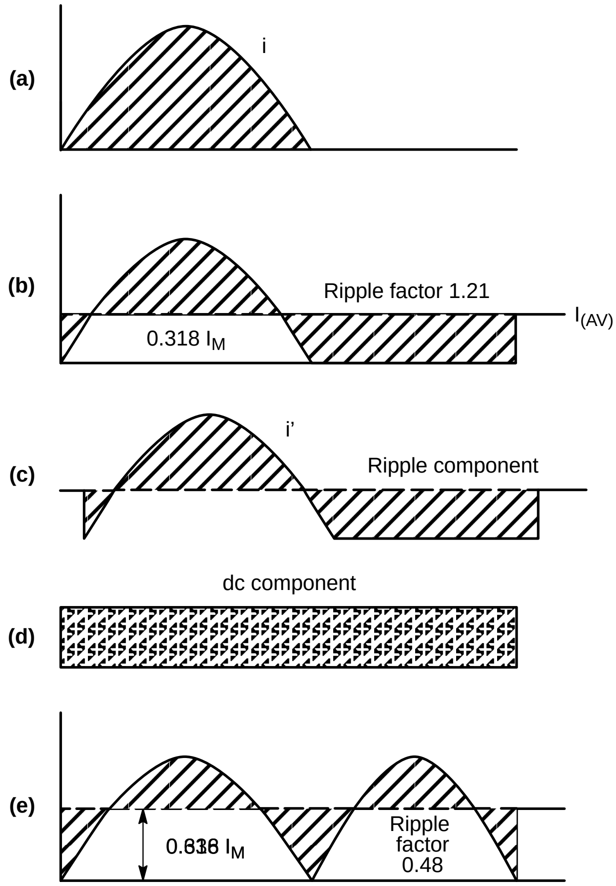


Figure 88. Illustration of Ripple Factors for Half–Wave and Full–Wave Circuits

The first term in this expression is the rms value of the total current $I_{(RMS)}$. The integral of the idθ part of the second term integrates to the average value $I_{(DC)}$, and the last term is simply $I^2_{(DC)}$ after the limits are applied. Thus:

$$I'_{rms} = \sqrt{I^2_{RMS} - 2I^2_{DC} + I^2_{DC}} \quad (5.10)$$

By combining Equations 5.10 and 5.5, the ripple factor may be expressed as

$$r_f = \sqrt{F^2 - 1} \quad (5.11)$$

(F is the form factor of the output current, not the diode current.) Substitution of values from Figures 86 and 87 into Equation 5.11 gives

$$\text{Half–wave } F = \frac{0.5 I_M}{0.318 I_M} = 1.57$$

$$r_f = \sqrt{1.57^2 - 1} = 1.21$$

$$\text{Full–wave } F = \frac{0.707 I_M}{0.636 I_M} = 1.11$$

$$r_f = \sqrt{1.11^2 - 1} = 0.482$$

Rectification Ratios

The preceding discussion of ripple factor leads to the idea that the heating losses (I^2R) which occur in the various parts of the complete rectifier circuit are increased by the irregular waveforms of current that are inherent in the rectifying process. This loss may be illustrated by comparing the heating loss caused by an ideal direct current to that resulting from the actual current waves of Figures 88(a), 88(b), and 88(c). The ratio of dc power in the load to ac rms power in the load is termed the rectification ratio, Ó and is written as

$$\sigma = \frac{I^2_{(DC)} R_L}{I^2_{(RMS)} R_L} \quad (5.12)$$

Substitution of the appropriate current values in terms of Im yields:

$$Ó = .406 \text{ (half–wave)},$$
$$Ó = .812 \text{ (full–wave)},$$

The rectification ratio is sometimes called the conversion efficiency. This term is somewhat misleading since the overall power efficiency for the assumed conditions (zero losses) must be 100%. The real significance of the rectification ratio is that it gives a qualitative indication of the increased heat losses that occur wherever a pulsating current flows through resistance elements. A second method of expressing the increased heat losses is by the current form factor, discussed earlier, which is the ratio of the root–mean–square to the average value.

Voltage Relationships

The input voltage required to achieve a given dc output voltage (V_{DC}) is a necessary piece of design information. The output voltage (V_{DC}) is I_{DC} R_L and Equation 5.2 states that I_{DC} = 0.318 I_M for the half–wave circuit. The peak current (I_M) is V_M/R_L and the rms input voltage $V_i = V_m/\sqrt{2}$

Combining these relationships yields:

$$V_i = 2.22\, V_{DC} \qquad \text{(half–wave)}$$

The full–wave circuit produces twice the dc level as the half–wave circuit for a given input voltage per transformer leg. Consequently,

$$V_i = 1.11\, V_{DC} \qquad \text{(full–wave)}$$

Another voltage of importance is the maximum voltage which the rectifier must block when it is not conducting. It is called the peak inverse voltage, or, in rectifier parlance, V_{RRM}. V_{RRM} is the sum of the peak input voltage and the rectifier peak output voltage at the same instant in time.

In the half–wave circuit of Figure 86, the output voltage is zero when the input voltage reaches its negative peak; therefore,

$$V_{RRM} = V_M = (\sqrt{2})(2.22\, V_{DC}) = 3.14\, V_{DC}. \qquad \text{(half–wave)}$$

In the full–wave center–tap circuit, the output voltage is maximum at V_M, because of rectifier conduction on one side, when the other side must block V_M. Therefore, $V_{RRM} = 2\ V_M$, but now the rms input voltage, V_M is 1.11 V_{DC}, consequently

$$V_{RRM} = (\sqrt{2})(2)(1.11)\ V_{DC} = 3.14\ V_{DC}.$$

(full–wave center–tap)

For the bridge, the blocking rectifiers are across the output voltage; therefore, the maximum reverse voltage is V_M. In this case,

$$V_{RMM} = (\sqrt{2})(1.11)\ V_{DC} = 1.57\ V_{DC}.$$

(full–wave bridge)

Circuit Variations

Many applications require equal positive and negative power supply voltages. Circuits shown in Figure 89 are commonly used to provide dual output voltages. These circuits are simply combinations of the standard half–wave and center–tapped full–wave circuits previously discussed. Chapter 2 illustrates how a rectifier bridge assembly may be properly handled thermally in the circuit of Figure 89(b). If the loads are balanced, the circuit of Figure 89(a) does not have a dc component in the transformer core, thereby overcoming the major objection to the half–wave circuit.

The standard rectifier diode, being just a two terminal device, has no internal means of regulating current. It is often used, however, in conjunction with controllable rectifiers such as SCRs in circuits similar to the hybrid bridge rectifier shown in Figure 90.

Similar circuits have a multitude of applications but are most often used with fixed input voltages to vary the average power to the load. A change in gate pulse delay to the 5CR serves to vary the conduction angle, α, as desired. The hybrid bridge may also be used to regulate the load power over a range of input voltages or perform a combination of these functions. In such circuits the form factor can be quite high. For this reason, additional derating must be applied to the rectifier diodes current rating when operation is confined to small angles of conduction.

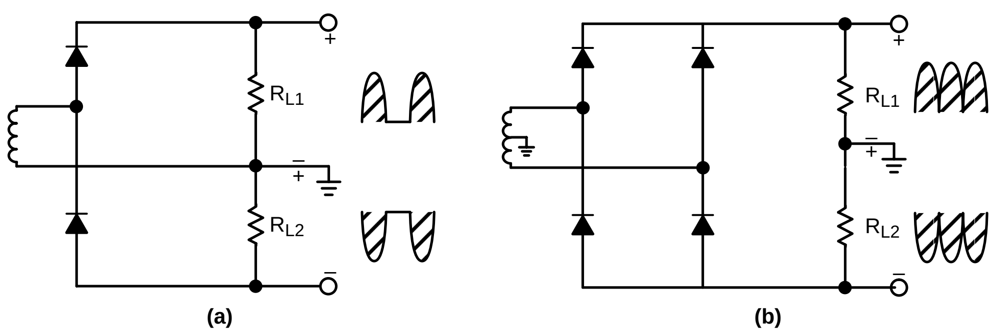


Figure 89. Single–Phase Supplies with Dual Output Voltages: (a) Dual Half–Wave Supply; (b) Dual Full–Wave Supply (a bridge assembly is convenient to use in this application)

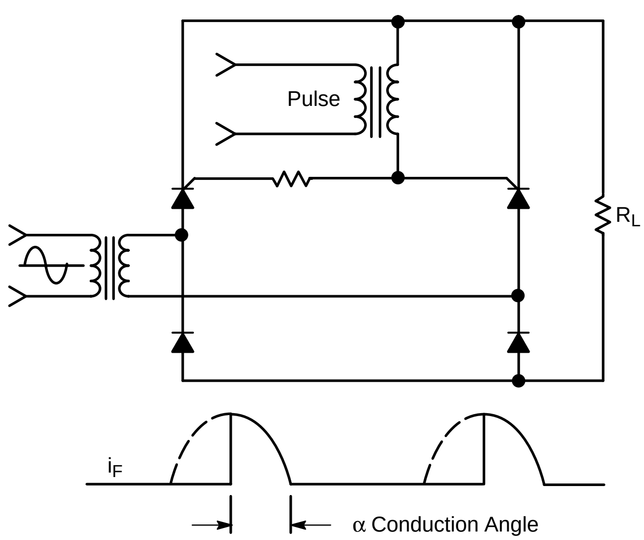


Figure 90. Typical Hybrid Bridge Rectifier with Resulting Diode Current Waveform

Summary

All the commonly used characteristics of low–frequency single–phase rectifier circuits with resistive loads have been discussed in this chapter. Results are summarized in Table 16. From the standpoint of maximum utilization of the transformer and reverse voltage rating of the rectifier diodes, the bridge circuit is superior to the other two. Although it requires more diodes, overall cost is generally similar because of simpler transformer construction and lower diode V_{RRM} requirements. At voltages of 12 V_{DC} and below, the power efficiency of the bridge becomes unacceptable because the output current must pass through two diodes in series. Furthermore, the diode V_{RRM} rating is of little concern. The full wave center–tap circuit is therefore preferable at low voltages. For economy reasons, the half–wave circuit finds use when current requirements are small, particularly if direct connection to the power line is allowable.

Filters are generally used on the output of single phase rectifying circuits. Circuit characteristics are similar to those shown in Table 16 when inductive loads or choke input filters are used, but characteristics change drastically with capacitive loads or filters. Chapter 7 discusses filtering and its effect upon rectifier circuit performance.

Chapter 6

Polyphase Rectifier Circuits

Polyphase Rectifier Circuits

A multiplicity of polyphase circuits have been devised, but only a few are of relative importance. This chapter develops some of the basic relationships to show the demands placed upon the rectifier diode by some of the more popularly used circuits. More exhaustive studies are given in the references [1] [2].

Polyphase transformer windings may be wye (Y) (also called “star”) or delta (Δ) connected. When Y connected as in Figure 91(a), voltages from any terminal (i.e.. A, B, or C) to the neutral (N) are called phase voltages and appear on a vector diagram as shown in Figure 91(b). The voltages V_{AB}, V_{BC}, and V_{CA} are called line voltages. The line voltages are also 120° out of phase, but because they are the vector sum of two phase voltages they are $\sqrt{3}$ times larger and are displaced 30° from the phase voltages as shown in Figure 91(c). A common method of transformer connection is the delta connection indicated in Figure 91(d). The voltage across the delta terminal is the same as the line voltage, and consequently bears the same relationship to the phase voltage as the line voltage.

By summing voltages from two phases in a transformer, a voltage bearing any desired amplitude and phase shift from a phase voltage may be obtained. Thus, there is no inherent limit to the number of phases which may be generated.

When a three–phase system is driving a balanced load, the power is given by

$$P = \sqrt{3}\,VI, \qquad (6.1)$$

where P = total power

V = rms phase or line voltage

I = corresponding rms phase or line current.

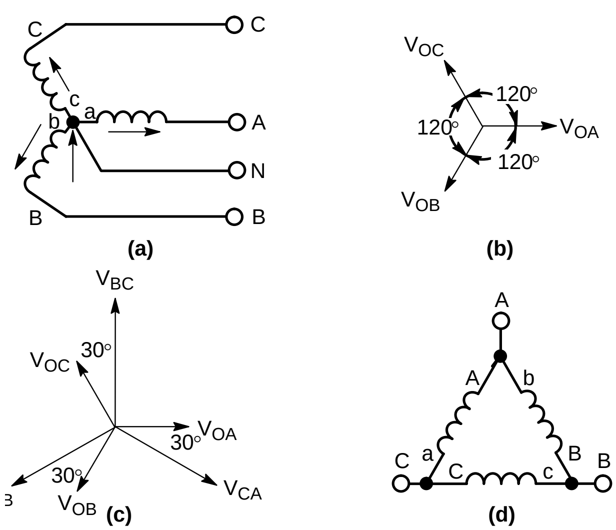


Figure 91. Y and Δ Connections and Appropriate Phase Relationships:
(a) Wye–Connected Transformer Windings;
(b) Phase Voltage Relationships;
(c) Complete Voltage Vector;
(d) Delta–Connected Transformer Windings

General Relationships in Polyphase Rectifiers

The analysis of polyphase rectifier circuits may be greatly simplified if the transformers and rectifiers are idealized; that is, they are assumed to possess no resistance or leakage reactance. Furthermore, the diode is assumed to have a zero forward voltage drop and a zero reverse current. By idealizing the components, simple general expressions can be derived for polyphase rectifiers.

Figure 92 shows the rectified voltage waveform for the general case of a polyphase rectifier circuit (See Figure 93 for the simplest polyphase rectifier circuit). Conduction takes place through the rectifier with the highest voltage across it. Note that neither the voltage nor the current is ever zero in the load. The instantaneous load voltage VL of Figure 92 is equal to the voltage of the conducting phase, and is given by

$$v_L = V_M \sin\theta \qquad (6.2)$$

where V_M = the peak phase–to–neutral voltage.

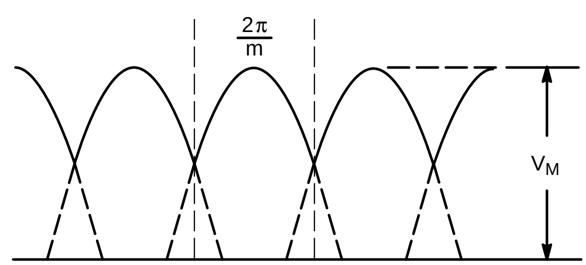


Figure 92. Rectified Voltage Waveforms for Multiphase Rectifier Circuit. The number of output pulses per cycle is designated

Current Relationships

As with single–phase circuits, the average or dc value of output current is of prime interest. It is the average value of the instantaneous rectified current over one cycle or a period of 2π/m radians (see Figure 92). Using integral calculus, the average current per phase equals the rectifier diode current and is given by

$$I_{F(AV)} = \frac{I_M}{2\pi}\int_{(\pi/2)+(\pi/m)}^{(\pi/2)+(\pi/m)} \sin\theta d\theta \qquad (6.3)$$

Solving,

$$I_{F(AV)} = \frac{I_M}{\pi}\left[\sin\left(\frac{\pi}{m}\right)\right] \qquad (6.4)$$

The total load current (I_L) is in times the current per conducting phase or

$$I_{L(DC)} = I_M\left(\frac{m}{\pi}\sin\frac{\pi}{m}\right) \qquad (6.5)$$

The rms current (I_f) is also of interest as it determines heating in the transformer winding resistance and in the diode. The rms value varies as the square of the instantaneous current and may be obtained by integration.

$$I_f = I_{F(RMS)} = \left[\frac{1}{2\pi}\int_{(\pi/2)+(\pi/m)}^{(\pi/2)+(\pi/m)} I_M^2 \sin^2 \theta d\theta\right]^{1/2} \quad (6.6)$$

Solving:

$$I_f = I_{F(RMS)} = I_M\left[\frac{1}{2\pi}\left(\frac{\pi}{m} + \frac{1}{2}\sin\frac{2\pi}{m}\right)\right]^{1/2} \quad (6.7)$$

It is convenient to have the diode rms current in terms of average current rather than the peak value so that the form factor of rectifier current may be determined. Equation 6.4 may be combined with Equation 6.7 to obtain

$$F = \frac{I_{F(RMS)}}{I_{F(AV)}} = \frac{\sqrt{\frac{\pi}{2}\left(\frac{\pi}{m} + \frac{1}{2}\sin\frac{2\pi}{m}\right)}}{\sin\frac{\pi}{m}} \quad (6.8)$$

For m greater than or equal to three, Equation 6.8 may be approximated as

$$F = \sqrt{m} \quad (6.9)$$

The simplification is done by using the small angle approximation, $\sin \theta = \theta$. Ordinarily, the error caused by the approximation would be too large for $\theta = 60°$, but because of the nature of Equation 6.8, the overall error is small. For example, for m = 3 Equation 6.8 yields F= 1.76 while Equation 6.9 yields F= 1.73.

Voltage Relationships

In order to design the transformer, the required input voltage from each transformer secondary leg (V_i) for a given output voltage must be known. Since $V_M = R_L I_M$ and $V_{L(DC)} = R_L I_{L(DC)}$, substituting these relationships into Equation 6.5 results in

$$V_{L(DC)} = V_M\left(\frac{m}{\pi}\sin\frac{\pi}{m}\right) \quad (6.10)$$

The rms value of V_i is $V_M\sqrt{2}/2$. Substituted and rearranging Equation 6.10 yields

$$V_i = \frac{V_{L(DC)}}{\sqrt{2}\left(\frac{m}{\pi}\sin\frac{\pi}{m}\right)} \quad (6.11)$$

The peak repetitive reverse voltage or peak inverse voltage applied to the rectifier, V_{RRM}, is also of vital interest. In full–wave center–tapped circuits, it is the sum of the instantaneous load voltage and peak input voltage.

To obtain exact values, the precise output waveform must be considered. However, V_{RRM} cannot exceed 2 V_M and in many polyphase circuits $V_{L(DC)}$ approaches V_M. To be conservative, let $V_L = V_M$; since $V_{i(PK)} = V_L\sqrt{2}$, using Equation 6.11 above, V_{RRM} can be put in terms of voltage as

$$V_{RRM} \le 2\left(\frac{V_{L(DC)}}{\left(\frac{m}{\pi}\sin\frac{\pi}{m}\right)}\right) \quad (6.12)$$

For bridge connections, V_{RRM} is simply the peak output voltage, which is one half of that given by Equation 6.12.

Transformer Utilization

High transformer utilization is one of the chief benefits of polyphase rectifier systems. The utilization factor (UF) is the ratio of do power in the load to the volt ampere or power rating of the transformer and is commonly used in comparing rectifier circuits. As defined,

$$UF = \frac{I_{(AV)}V_{(AV)(OUTPUT)}}{I_{(RMS)}V_{(RMS)(INPUT)}} \quad (6.13)$$

Values for the terms depend upon whether primary or secondary UF is desired. The utilization factor of the primary is found by considering the total volt–ampere requirement of the transformer and the total do load power. The secondary utilization factor generally considers only a particular winding's contribution to the load and the volt–ampere requirement of the winding.

Utilization factors can only be computed by studying the voltage and current relationships in a particular circuit configuration. It can be shown [1] that the highest obtainable utilization factor occurs when in = 3, i.e., conduction occurs in a winding for 120°. In this case, the winding is idle for two–thirds of the cycle. The larger the number of phases, the longer is the idle time and the less effectively the transformer is being used. To achieve high utilization factors and also the advantages of 6– or 12–phase operation, bridge circuits are used and special secondary winding arrangements have been devised which permit m to be three for the secondary windings; however, in is effectively six or twelve for the output voltage waveform.

Ripple

The same relationships for ripple hold true for polyphase circuits as with single–phase circuits. In Chapter 5, it is shown that the ripple factor of the output voltage may be expressed as

$$r_f = \frac{V_{mn}}{\sqrt{2}V_{L(DC)}} = \frac{\sqrt{2}}{(nm)^2 - 1} \quad (6.14)$$

An expression for the various ripple harmonics is obtained for the general polyphase rectifier system by applying Fourier analysis to the waveform. The resulting expression is given by

$$r_f = \frac{V_{mn}}{\sqrt{2}V_{L(DC)}} = \frac{\sqrt{2}}{(nm)^2 - 1} \quad (6.15)$$

where

r_f = ripple factor due to the nth harmonic,
V_{mn} = peak voltage of the nth harmonic,
$V_{L(DC)}$ = dc output voltage
m = number of output pulses per cycle,
n = order of the harmonic.

In Equation 6.15, the only harmonics involved are multiples of in. Thus, for m = 3, only the third, sixth, ninth, etc., harmonics of the input voltage contribute to the ripple voltage.

Rectification Ratio

The last general item of interest is the rectification ratio Ó. sometimes referred to as waveform or conversion efficiency. It is the ratio of dc power in the load ($I^2_{L(DC)}R_L$) to rms power in the load ($I^2_{L(RMS)}R_L$). In the ratio of terms, the load resistance cancels and $I_{L(RMS)}/I_{L(DC)}$ is the load current form factor. Consequently, the rectification ratio may be expressed as

$$\sigma = \frac{1}{[F(load)]^2} \quad (6.16)$$

To find the load current form factor, the total rms component of load current must be known. It may be found from Equation 6.6 by recognizing that the total current will be m times the phase current. Therefore.

$$I_{L(RMS)} = \left[\frac{m}{2\pi}\int_{(\pi/2)+(\pi/m)}^{(\pi/2)+(\pi/m)} I_M{}^2 \sin^2\theta d\theta\right]^{1/2} \quad (6.17)$$

Solving,

$$I_{L(RMS)} = I_M\left[\frac{m}{2}\left(\frac{\pi}{m} + \frac{1}{2}\sin\frac{2\pi}{m}\right)\right]^{1/2} \quad (6.18)$$

By comparing Equation 6.1810 Equation 6.7, it can be seen that rms load current is simply √m times rectifier leg current and the load current is in times the rectifier leg current. Therefore, the form factor of the load is simply 1/√m times the rectifier leg form factor as given by Equation 6.8. Accordingly:

$$F(load) = \frac{\sqrt{\frac{\pi}{2m}\left[\frac{\pi}{m} + \frac{1}{2}\sin\frac{2\pi}{m}\right]}}{\sin\pi/m} \quad (6.19)$$

An attempt to simplify Equation 6.19 as done with Equation 6.8 yields unity. Therefore, it must be used without approximation to yield meaningful results. Substituting Equation 6.19 into Equation 6.16 yields

$$\sigma = \frac{2m\sin^2(\pi/m)}{\left(\frac{\pi}{m} + \frac{1}{2}\sin\frac{2\pi}{m}\right)} \quad (6.20)$$

Overlap

In practice current does not switch instantaneously from one diode to another Overlap is observed, that is, a period of time where current flows in two diodes, each in different branches. Overlap is caused because leakage inductance and the stored charge in diodes will not allow current in the individual phases to change instantaneously.

One effect of overlap is to reduce the output of the rectifier circuit. This reduction, along with the voltage drop in the rectifier and transformer should be calculated and used to correct the output voltage predicted ideally. The overlap problem is discussed more fully in Chapter 7 and in the references.

Common Polyphase Rectifier Circuits

A discussion of the basic and more popular rectifier circuit configurations follows. Pertinent characteristics of each circuit are listed in Table 18.

Three–Phase Half–Wave Star Rectifier Circuit

The three–phase star rectifier circuit, often referred to as the three–phase half–wave rectifier, is illustrated in Figure 93(a). The associated voltage waveforms are shown in Figure 93(b). Because the circuit is economical, it finds limited use where dc output voltage requirements are relatively low and current requirements are too large for practical single–phase systems. The circuit is worth studying mainly because it is a building block of more complicated systems. By using the equations developed in the preceding section, the various items of interest may be calculated as given in Table 18.

The dc output voltage is approximately equal to the phase voltage. However, the diodes must block approximately the line–to–line voltage, which is √3 times the phase voltage. In addition, the transformer design and utilization are somewhat complicated in order to avoid transformer core saturation caused by the do component of current flow in each secondary winding.

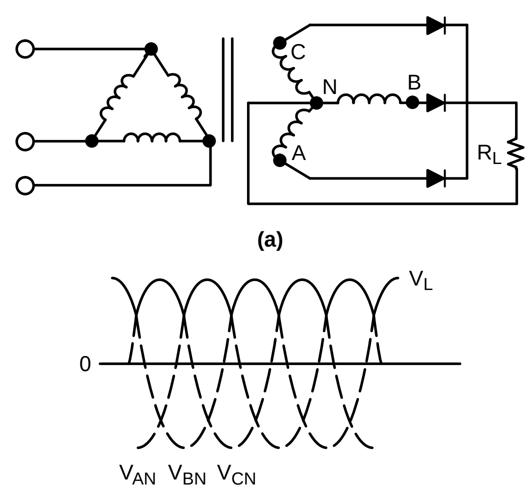


Figure 93. Circuit (a) and Waveform (b) for the Three–Phase Star or Half–Wave Rectifier Circuit

Three–Phase Inter–Star Rectifier Circuit

The three–phase inter–star or zig–zag rectifier circuit shown in Figure 94 overcomes some of the transformer limitations of the three–phase half–wave star circuit. The primary and secondary windings each consist of two coils, with pairs of coils forming a phase, located on different branches of the transformer core. The windings on the same core branch are connected such that the instantaneous magnetomotive force is zero. Although this connection eliminates the effects of core saturation and reduces the primary rating to the minimum of 1.05, it does so at the expense of economy, since it does not utilize the voltage of each winding to yield the highest possible out put voltage. The two secondary windings in series give a voltage of $\sqrt{3}V_{S1}$ instead of $2V_{S1}$. This results from the addition of two sinor voltages that are 600 apart. Consequently. the secondary volt ampere rating increases to 1.71 from the 1.48 value shown in Table 18. Otherwise, circuit constants are the same as the three–phase star circuit and therefore are not listed in Table 18.

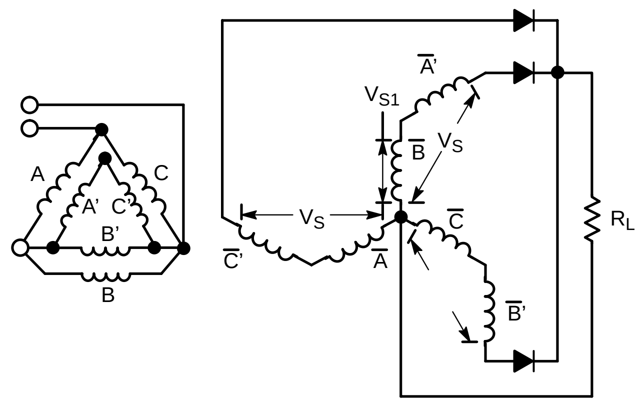


Figure 94. Three–Phase Inter–Star or Zig–Zag Circuit

Table 18. Summary of Significant Three–Phase Rectifier Circuit Characteristics for Resistive Loads

	Half–wave Star	Bridge	Double Wye with Interphase Transformer	Full–wave Star	Wye–Delta Connections	
					Parallel	Series
Average Current through Diode $I_{F(AV)}/I_{L(DC)}$	0.333	0.333	0.167	0.167	0.167	0.333
Peak Current through Diode $I_{FM}/I_{F(AV)}$	3.63	3.14	3.15	6.30	6.30	6.30
Form Factor of Current through Diode $I_{F(RMS)}/I_{(DC)}$	1.76	1.74	1.76	2.46	2.46	2.46
RMS Current through Diode $I_{F(RMS)}/I_{L(DC)}$	0.587	0.579	0.293	0.409	0.409	0.818
RMS Input Voltage Per Transformer Leg $V_i/V_{L(DC)}$	0.855	0.428	0.855	0.741	0.715	0.37
Diode Peak Inverse Voltage $V_{RRM}/V_{L(DC)}$	2.09	1.05	2.42	2.09	1.05	1.05
Transformer Primary Rating VA/P_{DC}	1.23	1.05	1.06	1.28	1.01	1.01
Transformer Secondary Rating VA/P_{DC}	1.50	1.05	1.49	1.81	1.05	1.05
Total RMS Ripple, %	18.2	4.2	4.2	4.2	1.0	1.0
Lowest Ripple Frequency, f_r/f_i	3	6	6	6	12	12
Rectification Ratio (Conversion Efficiency), %	96.8	99.8	99.8	99.8	100	100

1. See Table 19 in Chapter 7 for inductive load data.

Three–Phase Full–Wave Bridge Circuit

A three–phase full–wave bridge connection is commonly used whenever high dc power is required, as it exhibits a number of excellent attributes. It has a low ripple factor, low diode PIV, and the highest possible transformer utilization factor for a three–phase system. Because of the full–wave rectification associated with each secondary winding, it is permissible to use any combination of wye or delta primary and secondary windings or three single–phase transformers in place of one three–phase transformer. A schematic of a popular circuit is shown in Figure 95(a). The voltage waveforms are shown in Figure 95(b).

Each conduction path through the transformer and load passes through two rectifiers in series; a total of six rectifier elements are required. Commutation in the circuit takes place every 60°, or six limes per cycle. Such action is referred to as a six pulse rectifier which reduces the ripple to 4.2% and increases the fundamental frequency of ripple to six times the input frequency. No additional filtering is required in many applications. Thus, with this circuit the low ripple factor of a six–phase system is achieved while still obtaining the high utilization factor of a three–phase system. The dc output voltage is approximately equal to the peak line voltage or 2.4 times the rms phase voltage; each diode must block only the output voltage. Three–phase bridge connections are popular and are recommended wherever both dc voltage and current requirements are high.

The circuit characteristics are obtained by substituting m = 6 in the general equations. Results are shown in Table 18.

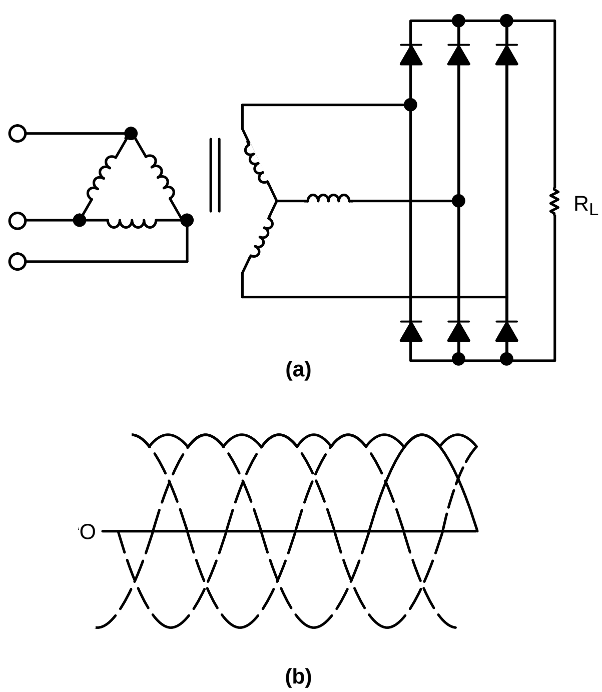


Figure 95. Three–Phases Full Bridge Circuit and Associated Waveforms.
(a) Three–Phases Full–Wave Bridge Circuit;
(b) Voltage Waveforms, Solid Line is Output Voltage and the Dashed Lines are the Phase Voltages

Three–Phase Double–Wye Rectifier with Interphase Transformer

The three–phase double–wye rectifier circuit is frequently used instead of a bridge circuit because each rectifier diode contributes only 1/6 instead of 1/3 of the load current. However, the peak inverse voltage for this circuit is higher than the three–phase star system due to the interphase reactor.

The circuit (Figure 96) consists essentially of two three–phase star circuits with their neutral points interconnected through an interphase transformer or reactor (also called a balance coil). The polarities of the corresponding secondary windings in the two parallel systems are reversed with respect to each other, so that the rectifier output voltage of one three–phase unit is at a minimum when the rectifier output voltage of the other unit is at a maximum, as shown. The action of the balance coil is to cause the actual voltage at the output terminals to be the average of the rectified voltages developed by the individual three–phase systems. The output voltage of the combination is therefore more nearly constant than that of a three–phase half–wave system; moreover, the ripple frequency of the output wave is now six times that of the supply frequency, instead of three times.

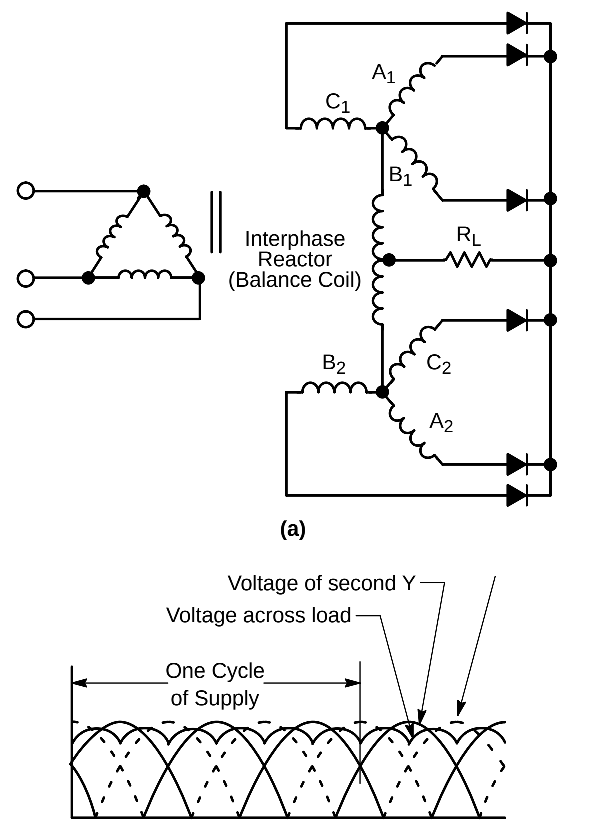


Figure 96. Circuit and Waveforms for the Three–Phase Double Wye Circuit: (a) Circuit; (b) Waveforms

In order that the individual three–phase half–wave systems may operate independently with current flowing through each diode one third of the time, the inter–phase reactor must have sufficient inductance so that the alternating current flowing in it as a result of the voltage existing across the coil has a peak value less than one–half the dc load current. That is, the peak alternating current in the interphase reactor must be less than the direct current flowing through one leg of the coil. Since the direct current flows in opposite directions in the two halves of the interphase reactor, no dc saturation is present in this reactor.

Six–Phase Star Rectifier Circuit

The six–phase star rectifier circuit shown in Figure 97 is often referred to as the three–phase diametric or full–wave rectifying circuit because it has a center–tapped transformer. The characteristics are obtained from the general rectifier equations where m is equal to six. Fields of application include requirements for very high dc load currents in low–to–medium voltage ranges. Voltage is usually restricted because the peak inverse voltage applied to the diodes is twice the peak phase voltage and transformer secondary utilization is poor. Current flows in only one rectifying element at a time, resulting in a low average current, but a high peak to average current ratio in the diodes.

The six–phase star circuit is attractive in applications which require a low ripple factor and a common cathode or anode connection for the rectifiers. The primary winding is generally delta–connected, although a wye connection is sometimes used with a tertiary winding. An additional advantage of the six–phase star is that the dc currents cancel in the secondary and, therefore, core saturation is not encountered.

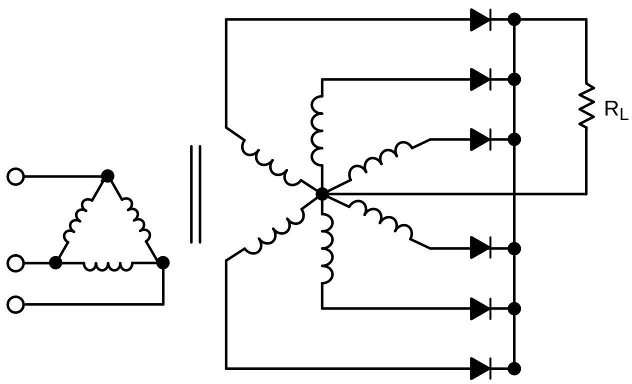
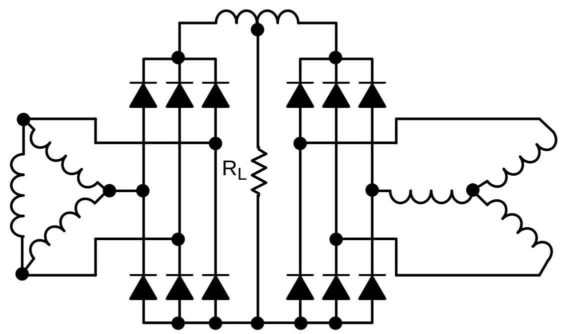


Figure 97. The Six–Phase Star Circuit

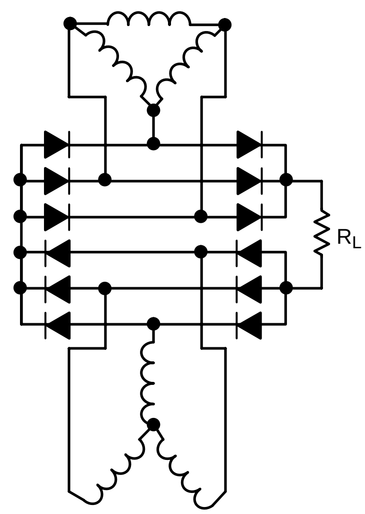
Six–Phase Full–Wave Bridge Circuits

In many applications, it is necessary to reduce ripple below the 4.2% level characteristics of tree–phase, full–wave and six–phase, half–wave connections. A reduction to approximately 1.0% at a ripple frequency twelve times the input frequency can be achieved by using wye–delta secondaries which result in 300 phase shift between windings. Either parallel or series bridge connections may be used for the output as shown in Figure 98.

When an equalizing reactor is used to couple the two bridge sections as in Figure 98(a), the system behaves as two parallel three–phase bridge circuits, with each section supplying one–half the load current. The parallel connection of the two groups is preferable to avoid having current pass through four diodes in series, but for high–voltage applications the series connection (as in part b) may result in a more economical design.



(a)



(b)

Figure 98. Six–Phase Full–Wave Bridge Circuits

Comparison of Circuits

As discussed, the three–phase and six–phase star circuits have limited applications because of a number of disadvantages. The most popular circuits by far are the three–phase full–wave bridge and the three–phase double–wye with interphase transformer. Both are alike in many respects.

Comparing the double–wye connection with interphase transformer with its rival, the three–phase bridge connection, it is found:

1. The diodes of one leg of an interphase transformer connection have to handle twice the voltage but only half the current of the diodes in one leg of a bridge connection with the same values of direct voltage and current. This is because the commutating groups operate in parallel in the interphase transformer connection and in series in the bridge connection. The high voltage rating is usually no problem and does not, in general, proportionally increase the price of the diodes (very high voltage excepted), but the lower current value reduces the rating of the diodes and all associated components such as fuses, bus bars, and cooling equipment, and is therefore a great advantage.
2. The total power losses are smaller in an interphase transformer connection because of the lower current carried by the diodes.
3. The reactive voltage drop caused by the bus bars can be made smaller in an inter–phase transformer connection since the current to be commutated is only half of the current commutated in a bridge connection, and the commutating voltage is twice as high.

For these reasons, the interphase transformer connection is competitive with the three–phase bridge connection in a certain voltage–current range, despite the relatively high transformer rating and the problem of avoiding saturation of the inter–phase transformer core due to current unbalance.

In special cases where low ripple is required and large power must be handled, a six–phase bridge or other high–order system may prove attractive. Other useful systems are covered in the literature [2].

High–Current Parallel–Connected Diodes

In some applications in chemical, metal, and transportation industries, the required dc output current cannot be achieved by using a single cell diode. In such instances, it is necessary to operate diodes in parallel. Care must be taken in the design to achieve a reasonable degree of current balance. The three main causes of unbalance are:

- differences in diode forward voltage drop
- self and mutual inductances of the diode and bus connection (referred to as the ladder effect in literature [3])
- magnetic coupling between legs.

Diode differences can be minimized by purchasing the diodes in closely matched sets but matched diodes are difficult to obtain from manufacturers at a cost–effective price. The resistance of the fuses used in series with each diode also improves balance. Some resistance will be encountered in the busing and connections. It should be very low in the main busses as compared to that in each individual diode leg.

Use of balancing transformers is a very effective means of forcing proper current balance and may be less expensive than purchasing factory matched units. The transformers consist of laminated iron cores usually with single–turn primary and secondary windings. The current from two diodes in parallel passes around the core in opposite directions so that any unbalance will induce a voltage which serves to correct the unbalance. The basic technique is shown in Figure 99. Its extension to larger numbers of rectifiers is illustrated and briefly discussed in Figure 100.

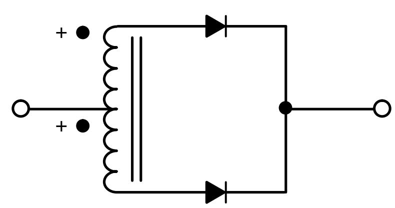


Figure 99. Parallel Operation of Rectifier Diodes Using Balancing Transformers

Computer simulation of the busing used in a stack of parallel diodes has led to the following design guidelines [4]:

1. The ac and dc buses of each leg should be as close as possible. Mutual inductance between these buses is beneficial.
2. The distance between the parallel diodes should be increased. The coupling of the mutual inductance of these paths acts as an improperly phased balance reactor.
3. The length of the diode path should be as short as possible. Although the resistance of the path increases as the path length increases, the self–inductance and mutual–inductance effects predominate, which increases the unbalance.

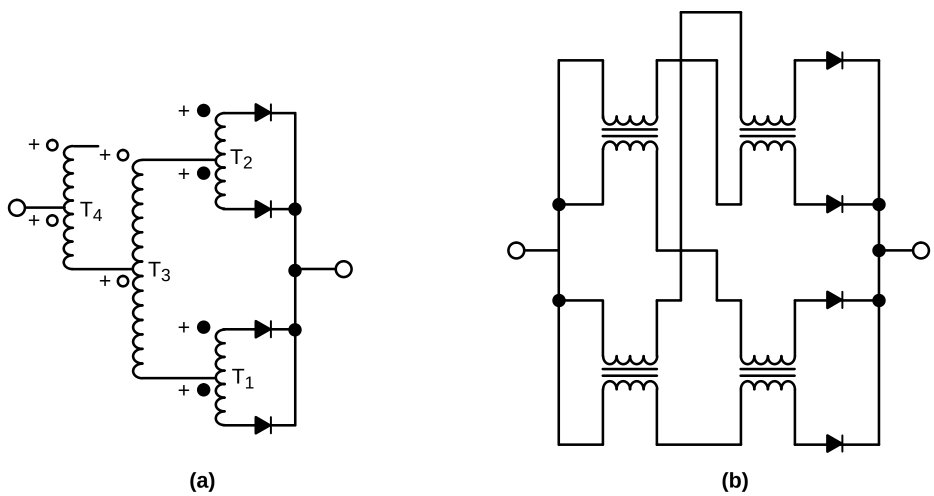


Figure 100. Schemes for Balancing the Current When More than Two Diodes are Required.
(a) Balancing Reactors Used with Diodes in an Extension of the Basic Circuit. Transformer T_3 must handle twice the current of T1 or T2; T4 must handle twice the current of T3, etc.;
(b) Balancing Reactors in a Closed–Chain System. Each winding carries only the current of one diode

References

1. Jacob Millman and Samuel Seely, *Electronics,* Second Edition, Chapter 14, McGraw–Hill Book Company, Inc., New York, New York, 1951.
2. Johannes Schaefer, *Rectifier Circuits: Theory and Design,* John Wiley and Sons, Inc., New York, New York, *1965.*
3. D.A. Paice, *Multiple Paralleling of Power Diodes,* IEEE Trans. TECI, Vol. IECI22, pp. 151–158, May, 1975.
4. S.M. Peeran, A. Kusko, W.R. Hodgson & E.G. Fellendort, Current Balance in Parallel Power Diodes in Three–Phase Rectifiers, IEEE Transactions on Industry Applications, Vol. IA–23, No 3., May/June, 1987.

Chapter 7

Rectifiers Filter Systems

Rectifier Filter Systems

Rectifiers without output filters, especially single–phase circuits, find limited application owing to their high ripple and relatively low rectification ratio. A Fourier analysis of the rectifier output waveform yields a constant term (the dc voltage) and a series of harmonic terms. Filters are usually added to extract the constant term and attenuate all harmonic terms. The input impedance of the filter dramatically influences the current and voltage relations of the transformer–rectifier combination; therefore, filters are broadly classified by the type of input element used.

Inductor–input filters are preferred in higher–power applications in order to avoid excessive turn–on and repetitive surge currents. Choke–input filters also offer greatly reduced electromagnetic interference, which is frequently caused by rectifier repetitive surge currents. More efficient transformer operation is also obtained because of the reduction in form factor of the rectifier current.

An inductor attempts to hold the load current constant and is thus more effective at heavy loads or small values of load resistance. Use of an inductor alone is generally impractical, particularly when variable loads must be handled because the attenuation is not sufficient with reasonable values of inductance. Analysis of a single L–R network shows that the ripple is reduced by the factor:

$$\frac{1}{\sqrt{1 + \left(X_L/R_L\right)^2}} \qquad (7.1)$$

where

X_L is the reactance of the filter choke

R_L is the resistance of the load.

Unless $X_L \gg R_L$, filter action is ineffective. The basic inductor–input filter is therefore an L section employing both inductance and capacitance so that the complementary characteristics of the filter elements are used to advantage. It should also be noted that inductive loads, or filters, are not good practice in half–wave single–phase rectifiers owing to the large rectifier inverse–voltage transient which occurs when conduction inevitably ceases.

Demands for smaller and lighter equipment have forced designers to use capacitor–input filters in lower–power systems operating from single–phase ac lines. When capacitor–input filters are used, diodes whose average rating more nearly matches the load requirement can be used if a source–to–load resistance ratio of about 0.03 and voltage regulation of about 10% are acceptable. Close regulation from the rectifier is seldom needed in electronic equipment because the rectifier is followed by an electronic regulator or switch mode dc–to–dc converter.

Behavior of Rectifiers with Choke–Input Filters

The voltage and current relations existing in rectifier systems having a series–inductance (choke–input) filter are illustrated by the sketches of Figures 101 and 102. While these apply to specific rectifier circuits, the same general behavior occurs in all rectifier systems of the choke–input type.

When the input inductance is assumed infinite, the current through the inductance is constant and is carried at any moment by the rectifier diode which has the most positive voltage applied to its anode at that instant. As the alternating voltage being rectified passes through zero or when another anode becomes most positive, the current suddenly transfers from one diode to another, giving square current waves through the individual rectifier diodes, as shown by the dotted lines in the lower sketches of Figure 101. When the input inductance is finite and large, the situation is as shown by the solid lines. The current through the input choke increases when the output voltage of the rectifier exceeds the average or dc value and decreases when the rectifier output voltage is less than the dc value, causing the current through the individual diodes to be modified as shown. If the input inductance is too small, the current decreases to zero during a portion of the time between the peaks of the rectifier output voltage, and the conditions are similar to a capacitor–input filter system, as discussed later in this chapter.

The ratio of peak current per individual diode to the dc current developed by the overall rectifier system depends upon the rectifier connection and the size of the input inductance. When the input inductance is infinite, the anode of each rectifier diode must at some time during the cycle carry the entire load current, except in the case of the three–phase half–wave double–wye system, where each diode is called upon to carry only half of the load current. These and other useful current relations for different rectifier connections are presented in Table 19. When the input inductance is not infinite, the current through the input inductance will vary around the value for the infinite inductance case, as illustrated in Figure 101. In the least favorable case, corresponding to an input inductance barely sufficient to obtain choke–input type of operation, the peak value of diode current is twice the value corresponding to infinite input inductance.

Rectifier Applications

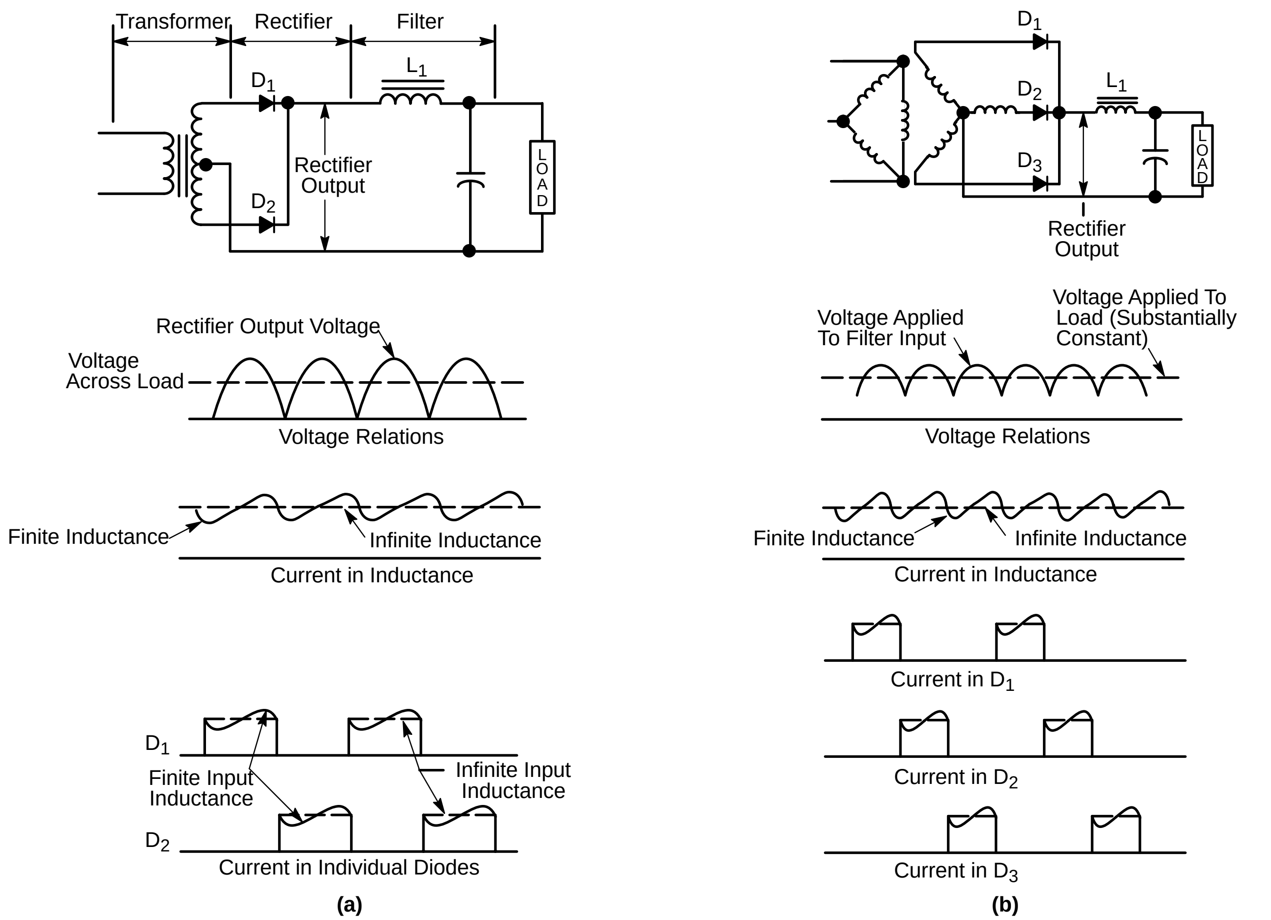


Figure 101. Voltage and Current Waveforms Existing in Rectifier Systems Operating with Choke–Input Filters: (a) Circuit of Rectifier and Filter, Single–Phase Full–Wave Case; (b) Circuit of Rectifier and Filter, Three–Phase Case

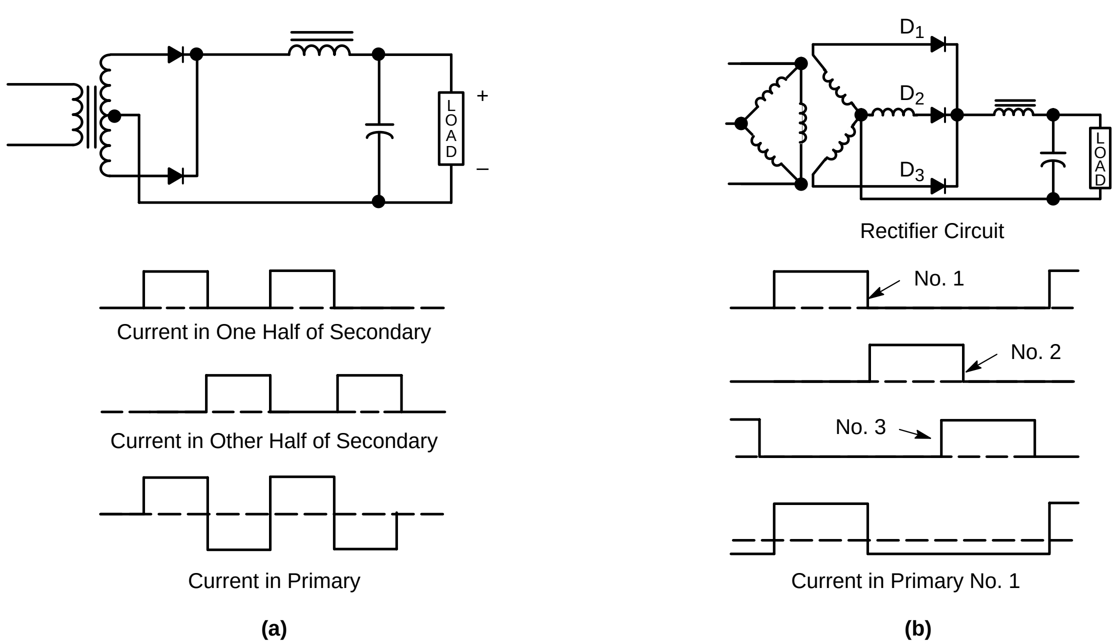


Figure 102. Typical Current Waves in Primary and Secondary Windings of Transformer for Ideal Choke–Input Systems: (a) Single–Phase Full–Wave System; (b) Three–Phase Half–Wave

Table 19. Characteristics of Typical Rectifiers Operated with Inductance–Input Filter Systems

Rectifier Circuit Connection / Characteristic	Single–Phase Full–Wave Center–Tap	Single–Phase Full–Wave Bridge	Three Phase Hall–Wave Star	Three–Phase Full–Wave Bridge	Three–Phase Double Wye With Interphase Transformer
‡Average Current Through Diode $I_{F(AV)}/I_{L(DC)}$	0.500	0.500	0.333	0.333	0.167
‡Peak Current Through Diode $I_{FM}/I_{F(AV)}$	2.00	2.00	3.00	3.00	3.00
Form Factor of Current Through Diode $I_{F(RMS)}/I_{F(AV)}$	1.41	1.41	1.73	1.73	1.76
RMS Input Voltage Per Transformer Leg $V_I/V_{L(DC)}$	1.11*	1.11	0.855	0.428	0.885
Diode Peak Inverse Voltage (PIV) V_{RRM} $V_{L(DC)}$	3.14	1.57	2.09	1.05	2.42
Transformer Primary Rating V_A/P_{DC}	1.11	1.11	1.21	1.05	1.05
Transformer Secondary Rating V_A/P_{DC}	1.57	1.11	1.48	1.05	1.48
Ripple ($V_F/V_{L(DC)}$ Lowest frequency in rectifier output (f_r/f_1) Peak Value of Ripple	2	2	3	6	6
Components: Ripple frequency (fundamental) Second harmonic Third harmonic	 0.667 0.133 0.057	 0.667 0.133 0.057	 0.250 0.057 0.025	 0.057 0.014 0.006	 0.057† 0.014 0.006
Ripple peaks with reference to dc axis: Positive peak Negative peak	0.363 0.837	0.363 0.637	0.209 0.395	0.0472 0.0930	0.0472 0.0930

1. This table assumes that the input inductance is sufficiently large to maintain the output current of the rectifier substantially constant, and neglects the effects of voltage drop in the rectifier and the transformers. $P_{DC} = 1^2{}_L R_L$, $V_L = I_L R_L$

*Secondary voltage on one side of center–tap.

†The principal component of voltage across the balance coil has a frequency of 3 fi and a peak amplitude of 0.500. The peak balance coil voltage, including the smaller, higher harmonics, is 0.605.

‡Assumes infinite input inductance.

The output voltage from the rectifier diodes applied to a choke input filter can, for nearly all practical purposes, be considered the same as if applied to a resistive load. The output voltage wave of the rectifier can be considered as consisting of a dc component upon which are superimposed ac voltages, termed ripple voltages. In the case of the idealized full–wave single–phase rectifier, Fourier analysis shows that the output wave v_L has the equation

$$V_L = \frac{2V_M}{\pi}\left(1 - \frac{2}{3}\cos\ 2\omega t - \frac{2}{15}\cos\ 4\omega t - \frac{2}{35}\cos\ 6\omega t\right) \tag{7.1}$$

where

V_M represents the peak value of the ac voltage applied to the rectifier diode

ω is the angular velocity ($2\ \pi f$) of the supply frequency.

(A full derivation of this equation is given at the end of this chapter.)

Note that the dc component of the output wave is $2/\pi$ times the crest value of the ac wave, and the lowest frequency component of ripple in the output is twice the supply frequency and has a magnitude that is two–thirds the dc component of the output voltage. The remaining ripple components are harmonics of this lowest frequency component and diminish in amplitude with the order of the harmonic involved in accordance with Equation 7.1. The more significant components of ripple are shown in Table 20. The ripple frequencies in the table are presented in relation to the fundamental frequency of ripple, not to the input frequency. For the full–wave circuit, the fundamental ripple component for a 60 Hz input is 120 Hz; the second harmonic is 240 Hz, and the third harmonic is 360 Hz.

Table 19 also gives the results of the analyses [1] for the output waves delivered by several popular polyphase rectifier connections. The ripple voltages are much less for the three–phase half–wave rectifier than for the single–phase connection, and are still less for the six–phase arrangements. In all cases, the amplitude of the ripple components diminishes rapidly as the order of the harmonics is increased. Not only are the ripple amplitudes lower but the ripple frequencies are higher for polyphase than for single–phase rectifiers, attributes which allow choke size to be greatly reduced. For example, the three–phase full–wave connections used with a 60 Hz input

have a lowest ripple frequency component of 360 Hz with an amplitude of 5.7% of the dc level. The second harmonic is 720 Hz with a level of only 1.4% and the third harmonic is 1080 Hz with an amplitude of only 0.6%. Consequently, a relatively small choke and capacitor can reduce the fundamental ripple frequency to a very low level and the harmonics to an insignificant level.

Input Inductance Requirements

To achieve normal choke–input operation, it is necessary to maintain continuous flow of current through the input inductance. The peak value of the alternating current component flowing through the input inductance must, hence, be less than the dc output current of the rectifier. The value for minimum inductance, called critical inductance (L_C) is derived from the following argument.

The average or dc current is $V_{L(DC)}/R_L$. The peak alternating current is very nearly the peak value of the fundamental–frequency component, since the higher–frequency components of the ripple current are relatively small. The current ripple harmonics are smaller than those given by Table 19, because of the higher reactance of the inductor at the harmonic frequencies. The inductor is normally followed by a capacitor having a low reactance at the fundamental ripple frequency, so that the peak ripple component is very close to $V_{M1}/\omega_1 L$. The ratio of peak–to–average current therefore closely approximates $(V_{M1}/V_{DC})/\omega L_C/R_L$, which must not be less than unity.

For a single–phase, full–wave circuit, the fundamental ripple component ω_i is twice the input frequency ω_i; also, the ratio V_{M1}/V_{DC} is the coefficient (2/3) of the fundamental ripple frequency term (cos 2ωt) in Equation 7.1. After substituting $2\omega_i$, for ω_1 and 2/3 for V_{M1}/V_{DC} into the previous expression, the following equation for critical inductance is found:

$$L_C = \frac{R_L}{3\omega_i} \qquad (7.2)$$

Using a similar procedure for polyphase rectifiers, the critical inductance is found to be

$$L = \frac{2R_L}{m(m^2 - 1)\omega_i} \qquad (7.3)$$

In Equations 7.2 and 7.3

L_C = the critical inductance in Henries
R_L = the load resistance in ohms
ω_i = the input or line frequency in rad/sec
m = the effective phase number for the output wave.

(If the resistance of the choke, diode, or transformer is significant, the values should be added to R_L).

For convenience, Equations 7.2 and 7.3 are put in the form

$$L_C = \frac{R_L}{A} \qquad (7.4)$$

where A is a constant determined from Figure 103.

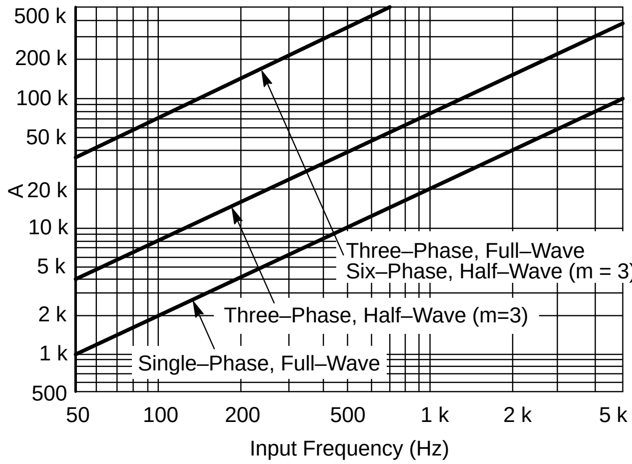


Figure 103. Values for the Constant A Used to Compute Critical Inductance

Values for other frequencies can easily be obtained by extrapolation, since A is inversely proportional to frequency. The value of A for a single–phase full–wave circuit operating at 20 kHz, for example, is 400,000 (10 times the 2 kHz value). The critical inductance required in polyphase systems is considerably less than that required in a single–phase system. The higher the load resistance (i.e., the lower the dc load current) the more difficult it is to maintain a continuous flow current. Also, with a given L, continuous flow will not occur when the load current drops below a critical value.

When the inductance is less than the critical value, the system acts similar to a capacitor–input system, described later. When the load current varies from time to time, it is necessary to satisfy the equations at all times if proper operation and, in particular, good voltage regulation are to be maintained. In order that this requirement may be satisfied at very small load currents without excessive inductance, it is usually necessary to place a resistance (commonly termed a bleeder resistance) across the output of the rectifier–filter system. The bleeder current is a compromise between having excessive dissipation in the bleeder resistor and excessively large L_c.

It is important to keep in mind that the effective inductance L_C is the incremental inductance, that is, the inductance to the alternating current superimposed upon the dc magnetization. Incremental inductance always increases as the dc magnetization decreases. This fact is of assistance in satisfying the equations at low load currents where R_L is large and can be put to advantage by the use of a “swinging” choke in which the inductance is varied by the load current using a controlled approach to saturation that lowers inductance as current increases.

It is essential to avoid series resonance between the input choke and first filter capacitor because at resonance very high circulating currents flow. Since at resonance, $1/\omega_1 L = \omega_1 C$, to avoid resonance it is sufficient to insure that

$$\omega_1^2 LC > 2 \quad (7.5)$$

where ω is the lowest frequency component of the ripple.

The output ripple magnitude depends upon the values of the choke and the capacitor used. It is calculated as shown in the subsequent sections.

Voltage Regulation in Choke Input Systems

Given ideal components, the load regulation of a rectifier–filter system employing input inductance greater than the critical value would be perfect–i.e., voltage output would be independent of load current. In practice, the output voltage falls off with increasing load as a result of resistance in the diodes, filter, and transformer and as a result of the leakage reactance of the supply transformer. The various resistances in the circuit reduce the output voltage without affecting the waveshapes in the system. The leakage reactance of the transformer, however, distorts the waveshape of the output voltage by preventing the current from shifting instantly from one transformer winding to another, as in the ideal cases of Figure 101. The situation in a typical case is shown in Figure 104 where μ represents the time interval required to transfer the current. During this transition period, the output voltage assumes a value intermediate between the open–circuit voltages of the two windings that are simultaneously carrying current, instead of following the open–circuit potential of the more positive anode. As a result, the average voltage of the output is less than if no leakage inductance were present by the amount indicated by the shaded areas in Figure 104(c). The quantitative relations, which are quite complicated, depend upon both the rectifier and the transformer connections.

When the input inductance is less than the critical value, the output voltage rises, and when the load resistance is very large, the output voltage will approach the peak value of the rectifier output waveform causing the system to have poor voltage regulation. When the load current is small, the energy stored in the inductor is also small and the inductor is essentially out of the circuit.

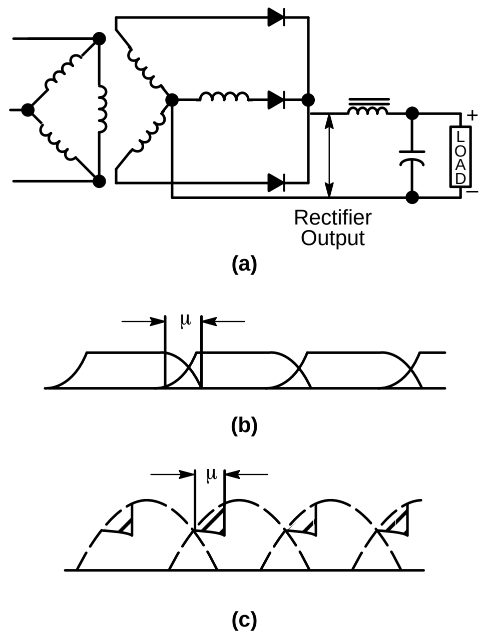


Figure 104. Effect of Transformer Leakage Inductance on the Behavior of a Polyphase Rectifier: (a) Three–Phase Half–Wave Rectifier Circuit; (b) Current of Individual Diodes; (c) Voltage Wave

Behavior of Rectifiers Used with Capacitor–Input Filters

The rectifier–filter system illustrated in Figure 105 differs from that of Figure 101 in that a shunt capacitor is presented to the rectifier output. Each time the positive peak alternating voltage is applied to one of the rectifier anodes, the input capacitor charges up to just slightly less than this peak voltage. No current is delivered to the filter until another anode approaches its peak positive potential. When the capacitor is not being charged, its voltage drops off nearly linearly with time because the load draws a substantially constant current. A typical set of voltage and current waveforms is shown in Figures 105(c) and 105(d). Use of an input capacitor increases the avenge voltage across the output terminals of the rectifier and reduces the amplitude of the ripple in the rectifier output voltage.

Analysis of Rectifiers with Capacitor–Input Systems

The detailed action that takes place in a capacitor–input system depends in a relatively complicated way upon the load resistance in the rectifier output, the input filter capacitance, the leakage reactance and resistance of the transformer, and the characteristics of the rectifier diode. For purposes of analysis, the actual rectifier circuit, such as that of Figure 105(a), is replaced by the equivalent circuit of Figure 105(b). The diode is replaced by switch S, which closes only when one of the diodes conducts current. The transformer is replaced by an equivalent generator having a voltage equal to the open–circuit line to center–tap secondary voltage and having equivalent internal impedance elements L_S and R_S. The inductance L_S is the leakage inductance of the transformer, measured across one–half the secondary winding with the primary short–circuited. The source resistance R_S the corresponding transformer's resistance plus a resistance to account for the resistance of the diode. In very low–voltage systems, a small potential ϕ may be needed to more accurately account for the diode voltage drop. The input capacitor of the filter system is C_1, and the first inductance L_1 (if used) and the load are assumed to draw a constant current I_1 equal to the dc voltage developed across the input capacitor divided by the sum of the actual load resistance plus the resistance of the filter inductance.

By utilizing the equivalent circuit of Figure 105(b), the effects that result from changes in individual circuit elements may be deduced. Thus, a decrease in the load resistance–i.e., an increase in the dc output current–reduces the average or dc output voltage, increases the ripple voltage, and increases the length of time during which the diode is conducting, as illustrated in Figure 106(a). Increasing the input capacitance has as its principal effect a decrease in the ripple voltage and also causes the average voltage to be increased slightly; these effects are shown in Figure 106(b). Increasing the leakage inductance (or resistance) of the transformer, reduces the average output voltage as illustrated in Figure 106(c), and likewise decreases the ratio of peak–to–average current flowing through the rectifier diode. In the case of a source and diode having very low impedance, the situation is as illustrated in Figure 106(d), and the peak current becomes quite high.

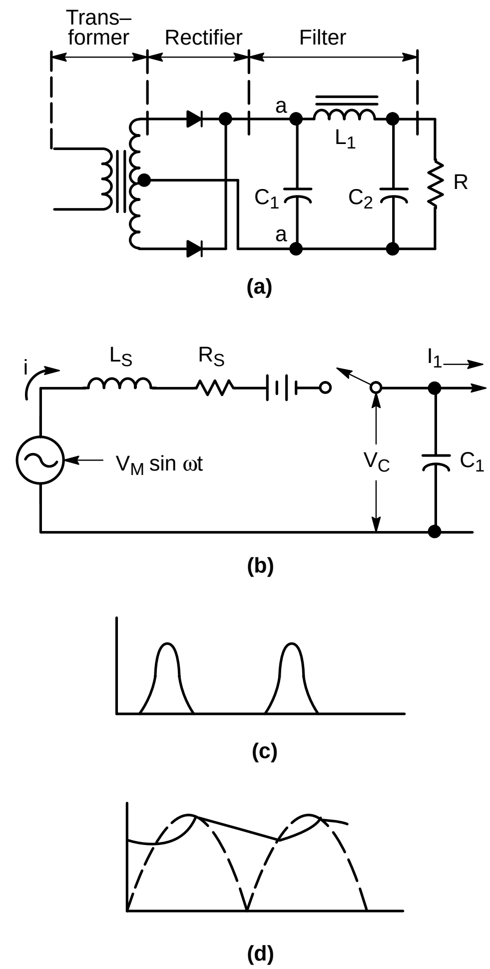


Figure 105. Actual and Equivalent Circuits of Capacitor–Input Rectifier System, Together with Oscillograms for Voltage and Current for a Typical Operating Condition:
(a) Actual Circuit; (b) Equivalent Circuit;
(c) Voltage Across Input Capacitor C1;
(d) Current Through Diode

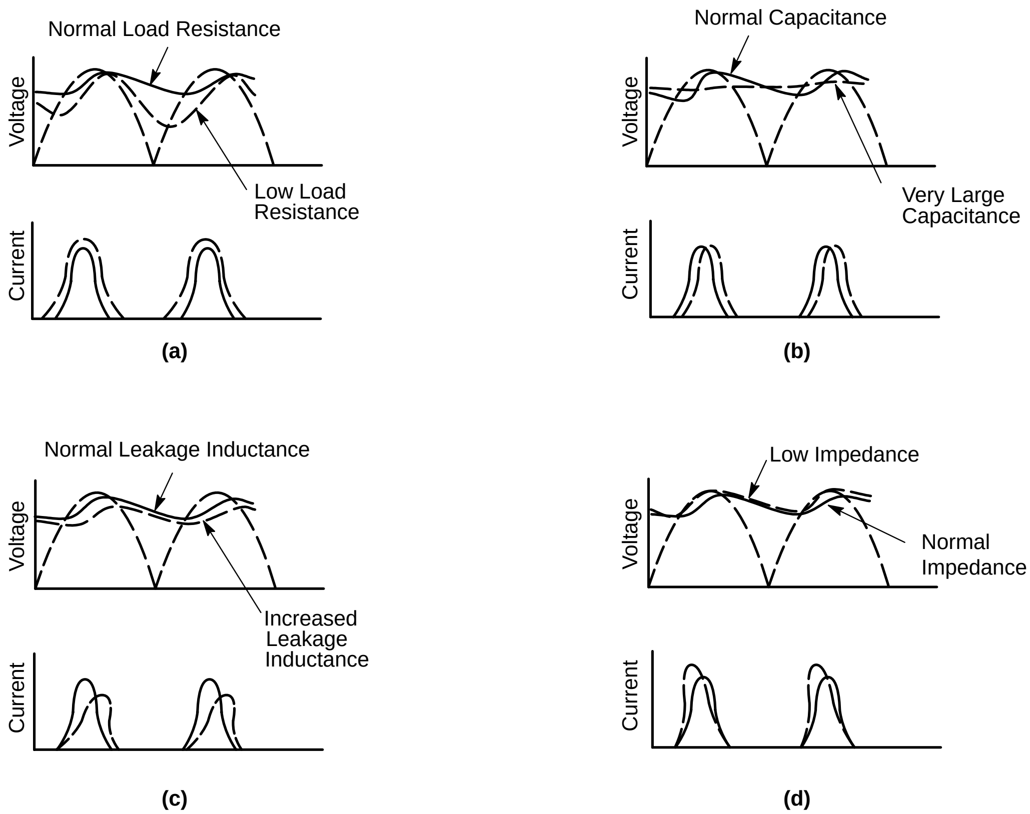


Figure 106. Effects of Circuit Constants and Operating Conditions on Behavior of Rectifier Having a Capacitor–Input Filter: (a) Effect of Load Resistance; (b) Effect of Capacitance; (c) Effect of Transformer Leakage Inductance; (d) Effect of Very Low Diode and Transformer Impedance

Design of Capacitor–Input Filters

The best practical procedure for the design of single–phase capacitor–input filters still remains based on the graphical data presented by Schade [2] in 1943. The curves shown in Figures 107 through 10 give all the required design information for half–wave and full–wave rectifier circuits. Whereas Schade originally also gave curves for the impedance of vacuum–tube rectifiers, the equivalent values for semiconductor diodes must be substituted. However, the rectifier forward voltage offset often assumes more significance than its dynamic resistance in low–voltage supply applications, as the dynamic resistance can generally be neglected when compared with the sum of the transformer secondary–winding resistance plus the reflected primary–winding resistance. The forward offset may be of considerable importance, however, since it is about 0.8 V for conventional silicon junction diodes and about 0.4 V for Schottky diodes, which clearly cannot be ignored in supplies for 12 V or less.

Rectifier Applications

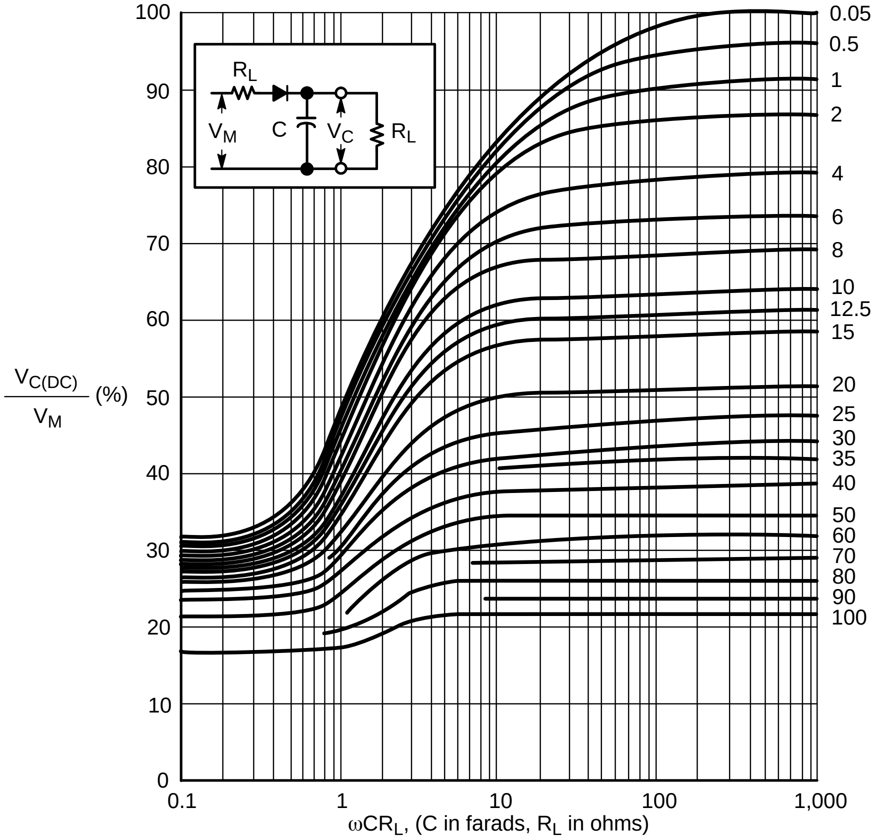


Figure 107. Relation of Applied Alternating Peak Voltage to Direct Output Voltage in Half–Wave Capacitor–Input Circuits (From O.H. Schade, Proc. IRE, Vol. 31, 1943, p. 356)

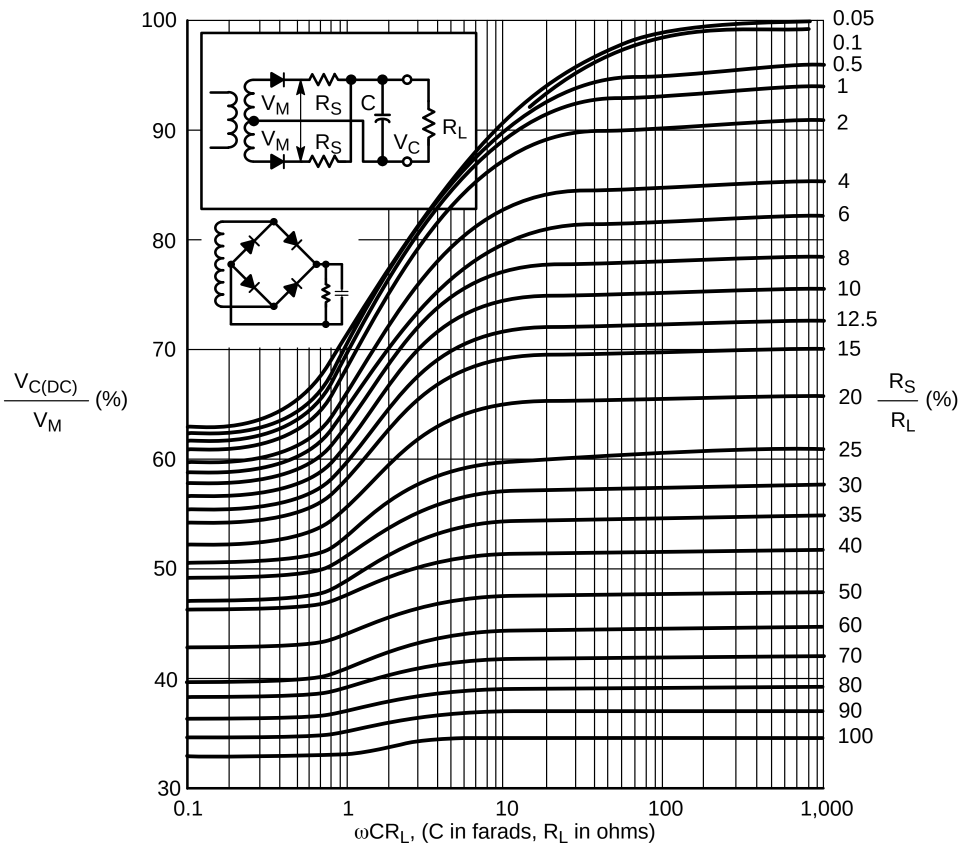
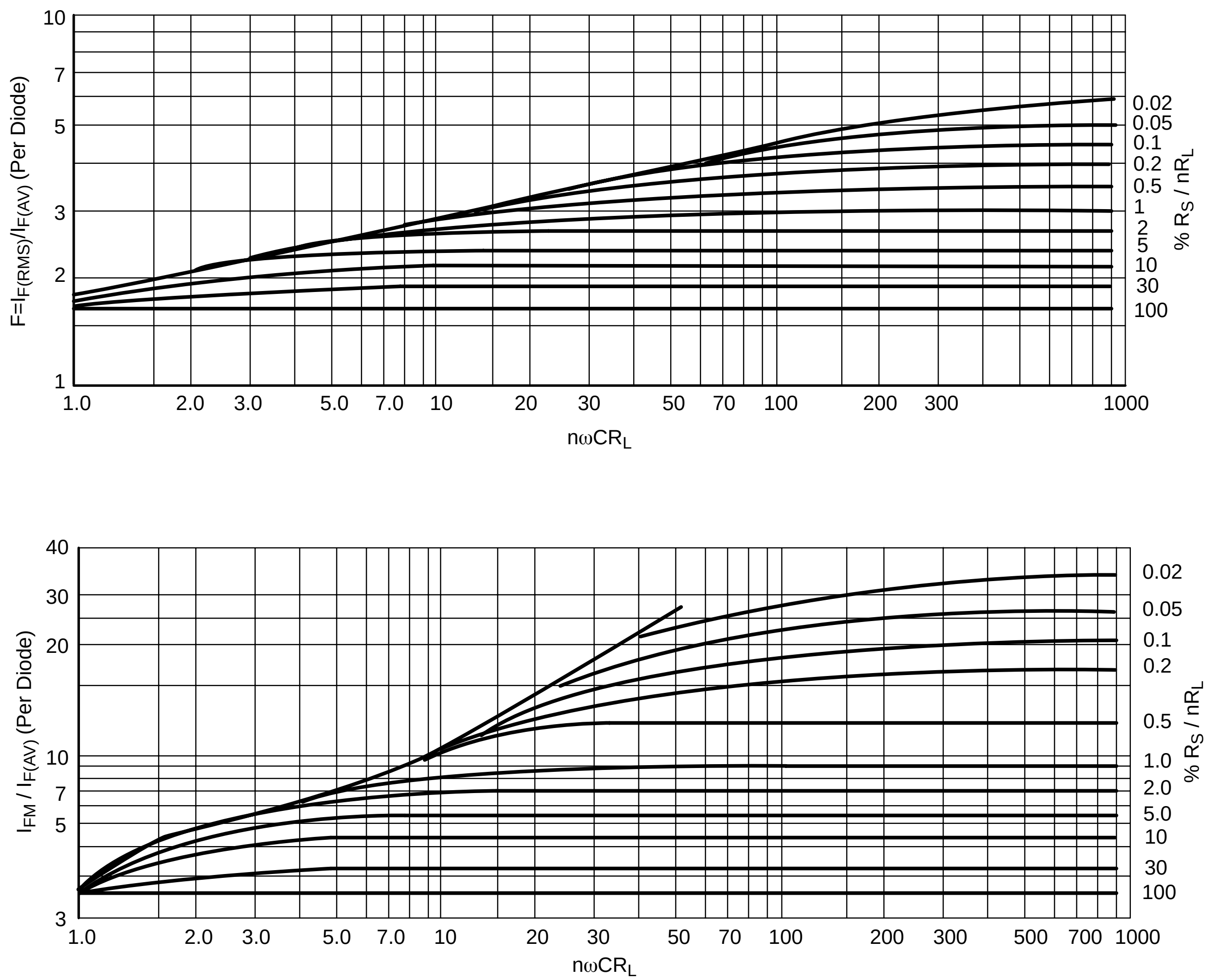


Figure 108. Relation of Applied Alternating Peak Voltage to Direct Output Voltage in Full–Wave Capacitor–Input Circuits (From O.H. Schade, Proc. IRE, Vol. 31, 1943, p. 356)

Figure 108 shows that a full–wave circuit must operate with $\omega CR_L \geq 10$ in order to hold the voltage reduction to less than 10 percent and $\omega CR_L \geq 40$ to obtain less than 2% reduction. However, these voltage–reduction figures require R_S/R_L, where R_S is now the total series resistance, to be about 0.1%. Such a low source resistance, if attainable, causes repetitive peak–to–average current ratios of 10 to 17 respectively, as can be seen from Figure 109. Rectifier diodes can handle high repetitive peak currents but their rated average current must be derated. To aid designers, manufacturers often provide capacitive load derating data.

The turn–on surge current generated when the input–filter capacitor is discharged and the transformer primary is energized at the peak of the input waveform causes a surge current determined by the peak secondary voltage less the rectifier forward drop limited only by the series resistance R_S. In order to control this turn–on surge, additional resistance must often be provided in series with each rectifier. It becomes evident that a compromise must be made between voltage reduction on the one hand, and diode surge rating and hence average current–carrying capacity on the other hand. If small voltage reduction–that is, good voltage regulation–is required, a much larger diode is necessary than that demanded by the average current rating. An obvious solution to this problem is to avoid the use of a capacitor–input filter. The alternative of a choke–input filter, as discussed previously, may be more costly than the cost of increased rectifier surge capacity.

A design example showing the use of Schade's curve is given at the end of this chapter. For comparative purposes, Table 21 shows rectifier circuit values when $\omega CR_L = 100$ and $R_S/R_L = 2\%$.



n = 1 For Half–Wave Single–Phase Rectifier Circuits
2 For Full–Wave Single–Phase Rectifier Circuits
ω = 2 πf, where f = Line Frequency

C in Farads
R_l in Ohms
R_S = RMD Equivalent Source Resistance

Figure 109. Relation of RMS and Peak to Average Diode Current in Capacitor–Input Circuits (From O.H. Schade, Proc. IRE, Vol. 31, 1943, p. 356)

Diode Peak Inverse Voltage

The peak inverse voltage applied to the diodes in a rectifier system operating in conjunction with a series–inductance input depends upon the topology. It ranges from as high as π times the dc component of the output voltage, in the case of the single–phase center–tapped connection, to barely more than the dc voltage in the three–phase full–wave bridge. Results in typical cases are given in Tables 19 and 21 and are the same as those determined for the resistive load circuits in Chapters 5 and 6.

The peak inverse voltage applied to the rectifiers in capacitor–input systems is generally based upon the capacitor charging to V_M which is the case when the load is light and the capacitor large. However, the peak of the reverse voltage sine wave occurs sometime after the capacitor has been charged. The peak voltage the capacitor attains is a function of the R_S/R_L ratio; the voltage on the capacitor when a diode's reverse voltage reaches its peak is dependent upon the load time constant. The relationship has been worked out for two cases as shown in Figure 111.

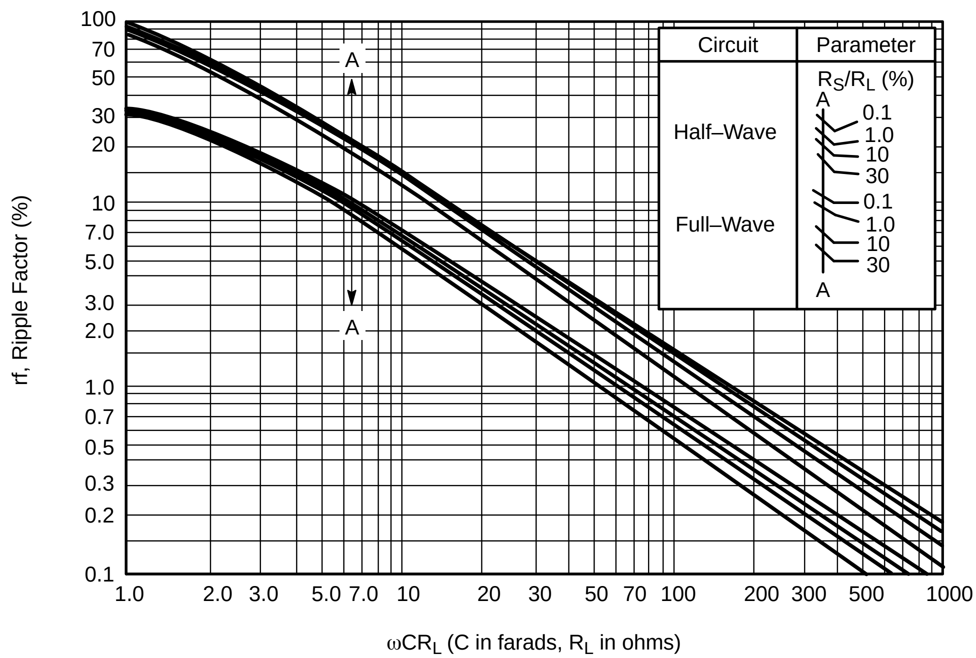


Figure 110. Root–Mean–Square Ripple Voltage for Capacitor–Input Circuits (From O.H. Schade, Proc. IRE, Vol. 31, 1943, p. 356)

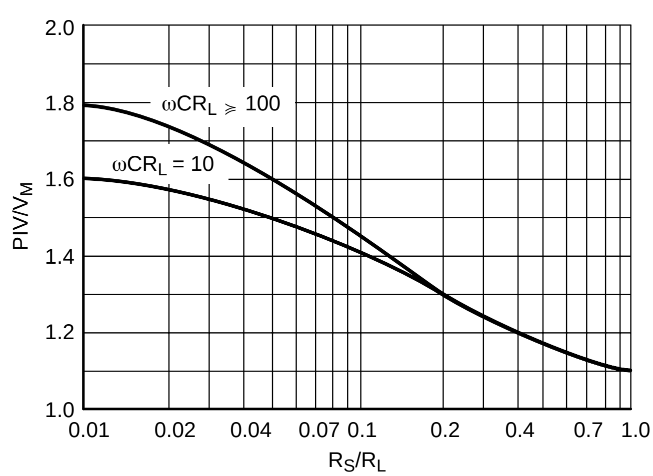


Figure 111. Ratio of Operating Peak Inverse Voltage to Peak Applied 60 Hz AC for Rectifiers Used in Capacitor–Input, Single–Phase, Filter Circuits

Transformer Considerations

The ratio of the actual dc rectified power output to the volt–ampere capacity of the windings on the basis of sinusoidal waves, for the same heat loss due to winding resistance, is termed the transformer utilization factor (UF). Its value depends upon the rectifier connections and is, in general, not the same for the primary and secondary windings, since the waveshapes in these windings will generally be different. Table 19 shows the reciprocal of the UF of the primary and secondary windings for some of the more commonly used rectifier connections. Derivation of the values is handled as shown for the resistive load circuits in Chapters 5 and 6.

The wave shapes of the currents that flow through the windings of the transformer in the idealized case of a rectifier system operating with an infinite input inductance are shown in Figure 102 for two typical cases. Because these waves are not sinusoidal, the heating of the transformer windings is greater for a given dc power output from the rectifier than would be the case if the same amount of ac power were delivered by the transformer to a resistance load. It is accordingly necessary to design the transformer windings more generously for rectifier applications than for cases where sinusoidal currents are involved.

Since transformer volt–ampere ratings vary with the rms content of the rectified circuit, they must be calculated for each capacitive–input filter design. Higher R_S/R_L ratios help reduce transformer requirements. Also, the effects of transformer saturation and finite bandwidth cannot be ignored. The former tends to limit the peak current while the latter allows the capacitor to charge over a longer time and hence reduces the peak charging current i = C dv/dt. Core saturation generally occurs on the recurrent charging peaks, and this will cause a lower output voltage than that calculated from the rated transformer secondary voltage. That saturation is occurring and that flattening is not due to the secondary–winding resistance can be determined by observing the voltage waveform induced in an unloaded winding on the same core.

Harmonics

Because the currents which flow in a rectifier circuit using a choke or capacitor input filter are not sinusoidal, high values of harmonic currents are present. These harmonic currents complicate the design of the transformer and, in single–phase circuits, the harmonic currents flow into the power line where they can cause a number of difficulties with the power distribution system. In a polyphase rectifier, the third harmonic can be kept from the powerline by having the primary windings arranged in a delta configuration. The circulating third harmonic component of primary current, of course, causes additional power loss in the transformer.

The choke input filter causes a square wave of current to flow in a single phase circuit. A Fourier analysis indicates that only odd harmonics are present and they decrease as 1/n, where n is the harmonic number. Thus, the amplitude of the third harmonic is 1/3, the fifth is 1/5, etc., that of the fundamental frequency.

Capacitive input filters cause a much higher current harmonic content than choke input filters cause. The harmonic content is particularly severe in transformer–less single phase supplies because transformer leakage inductance is not present. Inductance in series in the rectifier widens the diode conduction angle and lowers the current peak. A computer simulation of a single–phase bridge rectifier with a capacitor input filter yielded the results shown in Table 20 [3]. The value of the filter capacitor is normally so large that small changes in its value do not alter the results.

Table 20. Harmonic Currents In a Single–phase Full–wave Capacitor–Input Rectifier

Harmonic	Z_R*			
	1%	2%	5%	10%
3rd	83.0%	76.3%	63.5%	49.8%
5th	56.2%	42.4%	22.0%	10.2%
7th	29.2%	15.4%	8.1%	7.3%
9th	11.7%	8.7%	6.1%	3.8%
11th	8.6%	7.1%	3.5%	2.5%

*Z_R = Ratio of source inductive reactance to load impedance at the fundamental frequency.

The bridge is used without a transformer when its output is used by a switch mode dc–dc converter. In this case, the harmonic problems are transferred to the transformer in the electrical power distribution system feeding the equipment.

For a single–phase circuit, it is possible to significantly reduce harmonics fed into the power line by adding an optimally chosen inductor in series with the line to widen the conduction angle of diode forward current [4]. The inductor is much smaller than would be required for the typical continuous conduction choke–input filter. An optimally chosen capacitor added across the line raises the power factor to approximately 0.9, which is the same as obtained with a typical choke input filter.

Polyphase circuits behave differently. The larger the choke, the lower the harmonics and the higher the power factor. Increasing the inductor above a threshold does not offer any significant improvement in harmonics or power factor.

Surge Current

Because semiconductor junctions are limited in the amount of transient power that they can safely handle, attention must be given to the surge current occurring when the circuit is energized. The expected behavior can be deduced by studying the equivalent circuit of the rectifier system shown in Figure 105(b).

For analyzing a choke input filter, a choke is inserted between the switch and the capacitor C_1. Before voltage is applied, C_1 may be regarded as a short. Current buildup will be governed by the L/R_S time constant of the circuit. (L is the sum of the filter and transformer leakage inductance). The current surge is not significantly different than encountered

in normal operation, even if the input voltage is applied at a time when the voltage has reached its positive peaks, because the inductor has high reactance to fast rising transients.

The capacitor–input filter, however, allows a large surge to develop, because the reactance of the transformer leakage inductance is rather small. The maximum instantaneous surge current is approximately V_M/R_S and the capacitor charges with a time constant $\tau=R_SC_1$. As a rough–but conservative–check, the surge will not damage the diode if V_M/R_S is less than the diode I_{FSM} rating and τ is less than the time of one–half cycle. It is wise to make R_S as large as possible and not pursue tight voltage regulation; therefore, not only will the surge be reduced but rectifier and transformer ratings will more nearly approach the dc power requirements of the rectifier load.

Comparison of Capacitor–Input and Inductance–Input Systems

The basis for distinguishing between inductance–input and capacitor–input systems is that in the former the current flows continuously from the rectifier output into the filter systems, while in the latter the current flows intermittently from the rectifier into the filter. Intermittent action is also present with inductance–input systems when the input inductance is less than the critical value. When this is the case, the system is classified as a capacitor–input arrangement even though it possesses a series inductance.

A comparison of the performance of inductance–input and capacitor–input systems shows that, in the latter arrangement, the dc voltage is higher, the ripple voltage more, the surge current higher, and the voltage regulation poorer than when inductance input is used with the same diode, transformer, load resistance, and capacitance. Also, with capacitor–input the ripple voltage increases with increasing load current, unlike the inductance–input system where the ripple voltage is independent of load current. The utilization factor of the power transformer is much poorer with the capacitor–input system because of the higher ratio of peak–to–average current flowing through the rectifier diode, and likewise the diode is less efficiently utilized. The characteristics of several popular rectifier systems are shown in Table 21.

Shunt capacitor–input arrangements are generally employed in most consumer equipment such as personal computers, radio and television receivers, high–fidelity sound systems, or small public–address systems, when the amount of dc power required is small. They must be used with half–wave rectifiers and are attractive when the input power is taken directly from the ac line without a power transformer. In contrast, inductance–input arrangements are used when the amount of power required is large, since then the higher utilization factor and lower peak currents result in important savings in rectifier and transformer costs. Inductance input is often employed when good regulation of the dc voltage is important but the precision of an electronic regulator circuit is not needed. Inductance–input systems are normal in polyphase rectifier systems.

Filter Sections

Figure 114 gives typical examples of filters that are placed between the rectifier output and the load to make the current delivered to the load substantially pure direct current. These filters are made up of series impedances (either inductances or. resistances) that oppose the flow of alternating current from the rectifier output to the load and shunt capacitors that bypass the alternating currents that succeed in flowing through the series impedances.

For purposes of discussion and analysis, filters are ordinarily divided into sections, each consisting of a series impedance followed by a shunt capacitor, as indicated in Figure 114. In this classification, the inductance in an inductance–input system is considered to be part of the first section of the filter. However, a shunt capacitance across the rectifier output is not included as part of a filter section, but rather is considered to be part of the rectifier system, which delivers to the first filter section a voltage corresponding to the voltage developed by the rectifier across the input capacitor.

Table 21. Summary of Significant Rectifier Circuit Characteristics. Capacitive Data is for WCRL= 100 and RS/RL = 2.0%

Rectifier Circuit Connection		Single–Phase Half–Wave	Single–Phase Full–Wave Center–Tap	Single–Phase Full–Wave Bridge	Three–Phase Half–Wave Star	Three–Phase Bridge	Three–Phase Double–Wye With Interphase	Three–Phase Full–Wave Star
Characteristic	**Load**							
Average Current Through diode $I_{F(AV)}/I_{L(DC)}$	RL&C*	1.00	0.50	0.50	0.333	0.333	0.167	0.167
Peak Current	R	3.14	3.14	3.14	3.63	3.15	3.15	6.30
Through Diode	L*	–	2.00	2.00	3.00	3.00	3.00	6.00
$I_{FM}/I_{F(AV)}$	C	8.0	8.0	8.0	DATA NOT AVAILABLE			
Form Factor of	R	1.57	1.57	1.57	1.76	1.74	1.76	2.46
Current Through	L*	–	1.41	1.41	1.73	1.73	1.73	2.45
Diode $I_{F(RMS)}/I_{F(AV)}$	C	2.7	2.7	2.7	DATA NOT AVAILABLE			
RMS Current	R	1.57	0.785	0.785	0.587	0.579	0.293	0.409
Through Diode	L*	–	0.707	0.707	0.578	0.578	0.289	0.408
$I_{F(RMS)}/I_{L(DC)}$	C	2.7	1.35	1.35	DATA NOT AVAILABLE			
RMS Input Voltage	R&L	2.22	1.11	1.11	0.855	0.428	0.855	0.741
per Transformer Leg $V_i/V_{L(DC)}$	C	0.707	0.707	0.707	0.707	0.408	0.707	0.707
Diode Peak Inverse	R&L	3.14	3.14	1.57	2.09	1.05	2.42	2.09
Voltage PIV $V_{RRM}/V_{L(DC)}$	C	2.00	2.00	1.00	2.00	1.00	2.00	2.00
Transformer Primary	R	3.49	1.23	1.23	1.23	1.05	1.06	1.28
Rating V_A/P_{DC}	L	–	1.11	1.11	1.21	1.05	1.05	1.28
Transformer Secondary	R	3.49	1.75	1.23	1.50	1.05	1.49	1.81
Rating V_A/P_{DC}	L	–	1.57	1.11	1.48	1.05	1.48	1.81
Total RMS Ripple %	R	121	48.2	48.2	18.2	4.2	4.2	4.2
Lowest Ripple Frequency f_r/f_S		1	2	2	3	6	6	6
Rectification Ratio	R	40.6	81.2	81.2	96.8	99.8	99.8	99.8
(Conversion Efficiency) %	L*		100	100	100	100	100	100

*Inductive data valid when the circuit input voltage is a square wave, resistive or inductive load. $P_{DC} = I^2_L R_L$ (R_S neglected) $V_L = I_L R_L$

Voltage and Current Relations in Filters

The input voltage to the filter is whatever output voltage is developed by the rectifier connection in the case of inductance–input systems (it is the same as with a resistive load), or is the voltage developed across the shunt capacitor in the case of capacitor–input systems. The ripple voltage in the output of the rectifier–filter system is then the alternating component of this input voltage reduced by the action of the filter sections.

Most filter sections are composed of a series inductance and a shunt capacitance. In practical filters of this type, the reactance of the shunt capacitance at the lowest ripple frequency is much smaller than the resistance of the load (or the reactance of the series inductance of the following section). Substantially all ripple current entering the inductance L of the section flows through the capacitance C of the filter section. The current in the section is then $V_i/(\omega L - 1/\omega C)$, where V_i is the alternating voltage applied across the input to the filter section. The voltage that this current develops in flowing through the capacitor C is $(1/\omega C)V_i/\omega L - 1/\omega C) = V_i / \omega^2 LC - 1)$. Dividing by V_i results in the following relationship for the voltage reduction through an L–C section:

$$\frac{V_o}{V_i} = \frac{1}{\omega^2 LC - 1)} \quad (7.6)$$

where

L = series inductance of filter section

C = shunt capacitance of section

ω = angular frequency of ripple voltage involved.

Results of Equation 7.6 are given in Figure 113 for the usual case where the ripple components are harmonics of

60 Hz. Figure 115 indicates the actual ripple output of a full–wave single–phase rectifier circuit.

When the current drawn by the load impedance is small, the series inductance of the LC filter may be replaced by a series resistance as shown in Figures 114(e) and 114(f). This arrangement is widely used in low–current applications such as resistance–coupled amplifiers, tuned radio–frequency amplifiers, and so forth, and has the advantage that a resistance is much less expensive than an inductance of corresponding effectiveness. The disadvantage is the dc voltage drop and power loss that occur in the resistance; these losses limit the resistance–capacitance type of filter to cases where the current is small and where a moderate dc voltage drop is permissible.

In practical resistance capacitance filters, the reactance of the shunting capacitor is always made small compared to the series resistance of the filter and to the impedance to which the output of the filter section is connected. The ripple current flowing in R is $V_i\sqrt{R^2 + (1/\omega C)^2} \approx V_i/R$, where V_i is the ripple voltage applied to the section. The ripple output voltage of the section results from the current V_i/R flowing through the reactance $1/\omega C$ of the capacitance C of the section, and hence is $V_i/R\omega C$. Therefore, a section of resistance–capacitance filter reduces each frequency component of voltage applied to the input side according to the approximate relation:

$$\frac{V_o}{V_i} = \frac{1}{R\omega C} \qquad (7.7)$$

where

R is the series resistance

C is the shunt capacitance.

(See Figure 114(e)).

When a large amount of ripple reduction is required, more than one section is used because the total capacitance and inductance is less than that required for just one section. The total reduction in ripple voltage produced by the several sections is very nearly the product of the voltage–reduction factors of the individual sections. The effectiveness of the filtering accordingly increases rapidly with the number of sections. For many applications, a single section is entirely adequate, particularly with polyphase rectifiers. Only in special cases, such as audio–frequency amplifiers with very high gain, will the number of sections required exceed two, and then only for those parts of the amplifier that operate at very low signal–power levels.

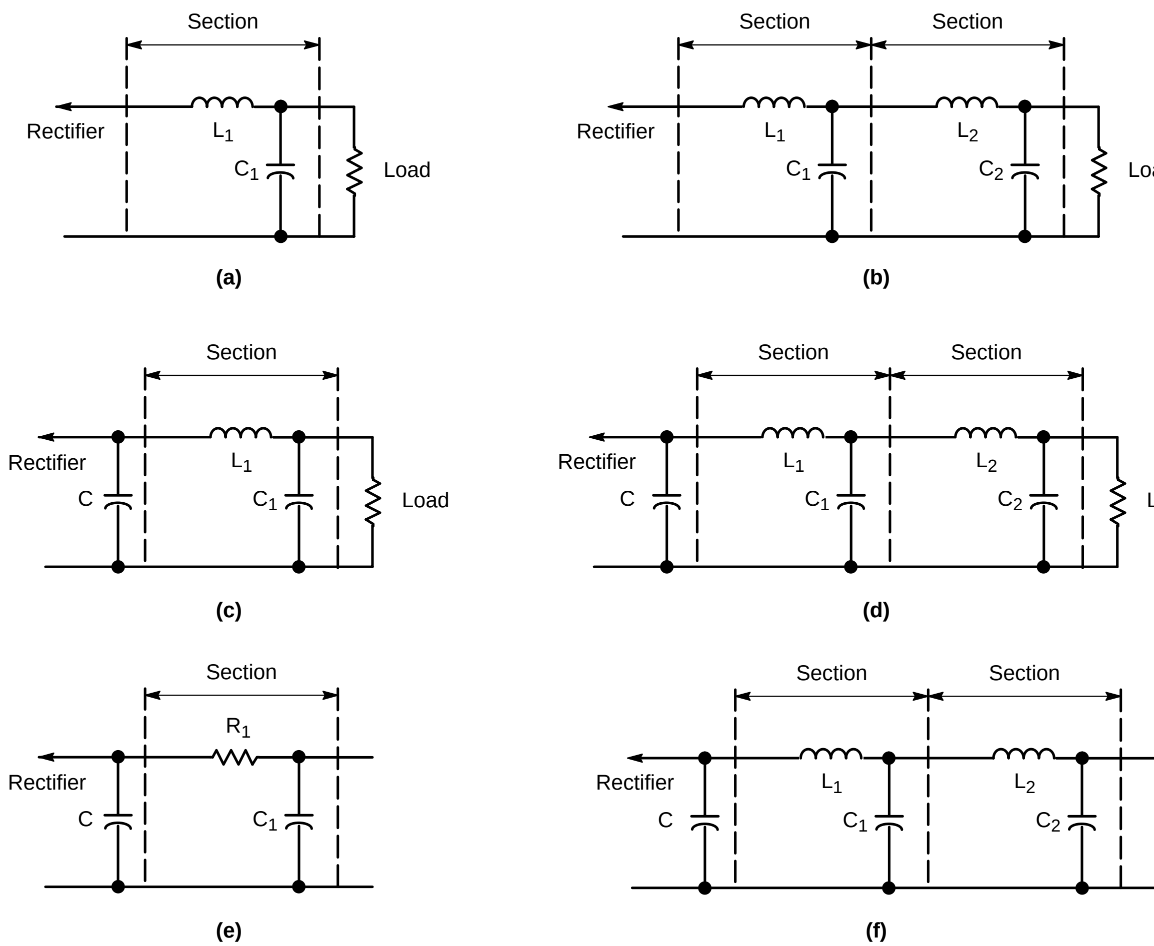


Figure 112. Typical L and π Section Filters: (a) and (b) L–C Filters, Choke Input; (c) and (d) L–C Filters, Capacitor Input; (e) and (f) Filters with Series Resistances

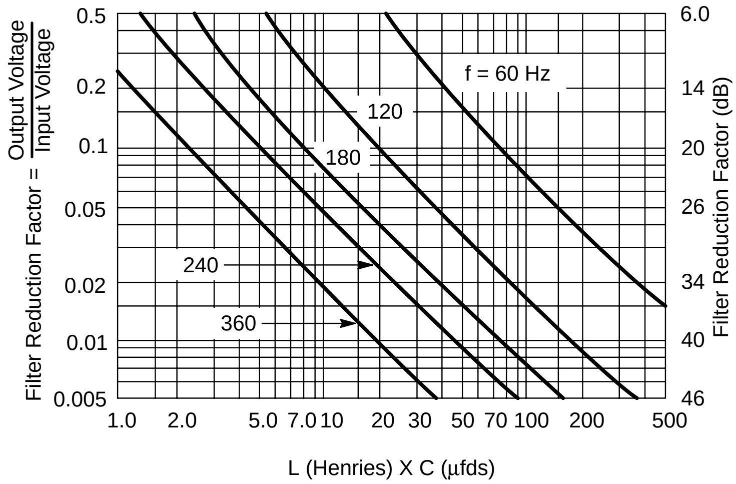


Figure 113. Reduction in Ripple Voltage Produced by a Single–Section Inductance–Capacitance Filter at Various Ripple Frequencies

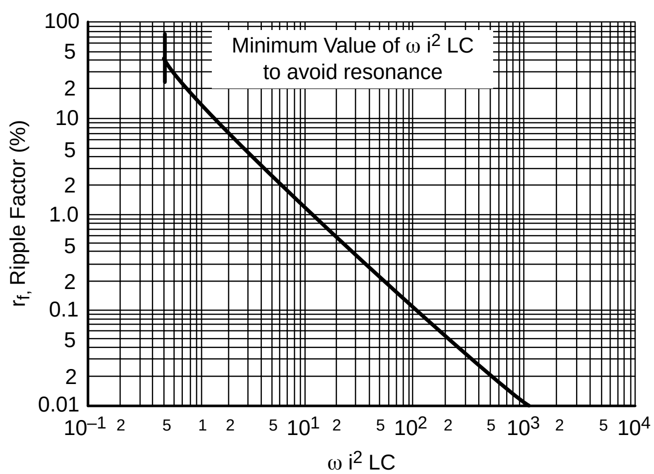


Figure 114. Ripple Factor for Single–Section *LC* Filter for Full–Wave Input

Graded Filters

When the output of a rectifier–filter system is called upon to supply voltages for several stages of an amplifier system, ordinarily the amount of ripple or hum voltage that can be tolerated is least for those stages that operate at the lowest signal–power levels. This makes it desirable to arrange the filter system so that the voltages applied to different circuits operated from the rectifier–filter system undergo different amounts of filtering.

For example, the output stage of an audio amplifier obtains its collector voltage directly from the input capacitor, which is permissible because of the high power level at which the output stage operates, combined with the hum–suppressing action of the usually used push–pull connection. Progressively increased filtering, is provided for the lower level stages, care being taken to design the system so that the reduction in ripple voltage introduced by the filter between stages is at least as great as the amplification of the stages.

A graded filter reduces to the lowest possible value the magnitude of the currents that must be carried by the series impedance arms of the filter and often makes it practical to use resistance–capacitance filter stages in parts of the system, as illustrated in Figure 114. This results in substantial economy in cost, weight, and size compared to an arrangement in which the entire rectifier output is subject to the maximum amount of filtering. A graded filter also provides isolation or decoupling between amplifier stages, thereby reducing regeneration.

Filter Component Selection

The inductors used in a low–frequency filter must have laminated iron cores with an air gap that is sufficient to prevent the dc magnetization from saturating the core. The inductance that is effective in the filter is the incremental inductance, which depends both upon the dc and the ac magnetizations of the core. In estimating the ac magnetization that can be expected, it is normally assumed

that the alternating current flowing in the inductance is equal to the voltage of the lowest ripple frequency applied across the input of the filter section, divided by the reactance of the inductance of the section. The alternating magnetization in the inductance of the first section may be relatively large, whereas the alternating magnetization for the inductances of the other filter sections will be very small if the first section is at all effective.

The capacitors used in filters must be capable of continuously withstanding a dc voltage equal to the peak voltage applied to the rectifier. Electrolytic capacitors are ordinarily used where the peak voltages do not exceed 400 to 500 V. Such capacitors have very low cost in proportion to capacitance, but they possess the disadvantage of a limited life. Various types of plastic capacitors are usually used at higher voltages and also find use at lower voltages where long life is more important than low cost.

Example of Power Supply Design

Information given in this chapter is used to design a power supply for an audio amplifier in a home entertainment center. The dc supply voltage will be obtained from a 120 V, single–phase line. Electronic regulation is not necessary, economically feasible, or particularly desirable for such applications. The supply will consist of a step–down transformer, rectifier diodes, and a capacitive–input filter as shown in Figure 116.

Design Constraints

For this example, a supply is to be designed to meet the power requirements of a 20 W stereo amplifier (operating in a maximum ambient temperature of 55°C) which must drive 8–Ω speaker loads. The output power transistors are in a totem–pole configuration and operate from a single dc supply. Theoretically, a 36 V supply would be adequate, but a full load voltage of 40 V is required to make up for transistor and resistive losses. The average current drawn from both channels at rated power out (20 W per channel) is 1.43 A. To minimize distortion, idle current is 50 mA per channel. At zero signal, the ripple is critical and must be held to 400 mV maximum so that no hum is detectable. Regulation is not critical, but it is desirable to keep the no–load voltage below 50 V to ease the voltage requirements of the output transistors.

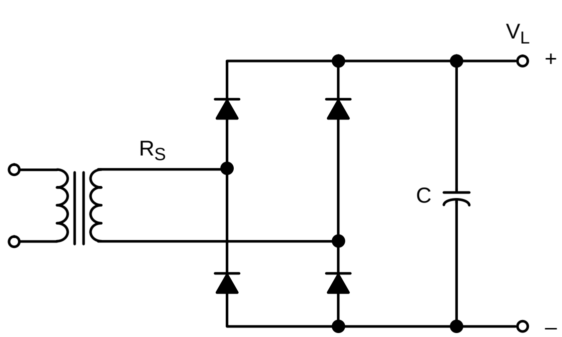


Figure 115. Bridge Circuit Used for Example Design

From the previous discussion, the power supply specifications are summarized in terms of the constants required:

1. $\omega = 2\pi f = 377$ rad/sec
2. $R_L = V_{L(DC)}/I_{L(DC)}$
   $= 40/1.43 = 28\ \Omega$ at full load
   $= 50/0.1 = 500\ \Omega$ at idle
3. $V_{L(DC)}$(full load)/$V_{L(DC)}$(idle) = 80%
4. r_f = 400 mV/50 V = 0.8%

Passive Component Selection

The choice of circuit is between a single–phase full–wave center–tap or a full–wave bridge circuit as outlined in Table 16 in Chapter 5. The center–tap requires only two diodes, but it has the disadvantages of requiring an additional winding, causing a poorer transformer utilization factor, and doubling the required voltage ratings of the diodes. The deciding factor in favor of using the bridge rectifier circuit is its higher secondary utilization factor. This is an important consideration when using a capacitive input filter because high rms currents are required from the transformer to obtain good output voltage regulation (i.e., a low R_S/R_L ratio is required).

The curves shown in Figure 110 are used to find a value for C to meet the ripple specification. To use the curves, it is necessary to estimate a value for the R_S/R_L ratio. This can be done by referring to Figure 108 and assuming that the idle dc voltage equals the peak of the input voltage and that the ωCR_L product is large enough to place operation in the right hand plateau of Figure 108. Based on these assumptions, note that to hold $V_{L(DC)}/V_M$ above 80%, R_S/R_L < 7%. If it is desired to minimize[1] C, then operation may move into the knee region and R_S/R_L may need to be made considerably less than 7% to hold $V_{L(DC)}/V_M$ to 80%.

From Figure 110, at ω_f = 0.8%, R_S/R_L = 1% (being conservative), read $\omega CR_L = 90$. Therefore, C = 90/(377)(500) = 496 µF and a standard value of 500 µF can be used.

A suitable value for R_S to maintain the required voltage regulation can be obtained by using Figure 108. Under full load with C = 500 µF, $\omega CR_L = 5.2$. Again assuming that the output voltage at idle equals the peak input voltage, which makes $V_{L(DC)}/V_M$ = 80%, read $R_S/R_L \approx 3.5\%$ at $\omega CR_L = 5.2$. Therefore, $R_S = (0.035)(28) \approx 1.0\ \Omega$. This value may be largely composed of the sum of the transformer secondary resistance, the reflected primary resistance, and the diode dynamic resistance.

At idle, R_S/R_L = 1.0/500 = 0.2% and, as previously found, $\omega CR_L = 90$. From Figure 108, $V_{L(DC)}/V_M$ is read as 97%, which verifies the validity of the assumption that at idle $V_{L(DC)} = V_M$.

[1]In many cases, minimizing C will result in lower cost. However, by choosing a larger value of C, R_S may be increased to obtain the same voltage regulation, with the beneficial effects of lower ruts and peak surge currents. The lower rms values may result in lower cost for the transformer and diodes, so that the overall power supply cost is less.

Transformer Selection

The peak output voltage developed across the capacitive filter has been specified as 50 V, which requires a transformer secondary rms voltage of 35 V. However, because of the forward voltage drops across the diodes, a standard 36 V transformer may be used. The transformer volt–ampere rating is dependent on rms current, which can be obtained from Figure 109. At full load $R_S/R_L = 3.5\%$ and $\omega CR_L = 5.2$. In full–wave circuits, n = 2, the value of $n\omega CR_L = 10.4$, and $R_S/R_L = 3.5\%$, yielding a diode current form factor of approximately 2.1. The load and secondary form factor, being full–wave, is $1/\sqrt{2}$ times the diode form factor. Consequently, the transformer rating is

$$VA = (36)\,(1.43)\,(2.1)/\sqrt{2} = 77.$$

Diode Selection

A diode bridge assembly is chosen for ease of installation. Since each diode's reverse voltage is clamped by the filter capacitor, the bridge voltage rating need only be sufficient to handle a high line condition, typically 130% of nominal. In this application, 100 V diodes are adequate.

In choosing a bridge, the total output current of 1.43 A is used as a selection guide rather than average current per diode. A series of bridges with 2.0 A current ratings will be evaluated for this application. It is necessary to determine whether these bridges can handle the repetitive peak currents into the filter and the initial surge current required to charge the filter capacitor.

The steady–state peak–to–average current ratio can be obtained from Figure 109. With $R_S/R_L = 3.5\%$ and $n\omega CR = 10.4$, a peak–to–average ratio per leg of 7 is interpolated. Consulting Figure 116, taken from the diode's data sheet, a peak–to–average ratio of 7 per leg yields a maximum full–wave average current capability of about 1.7 A at 55°C, which exceeds the requirement comfortably.

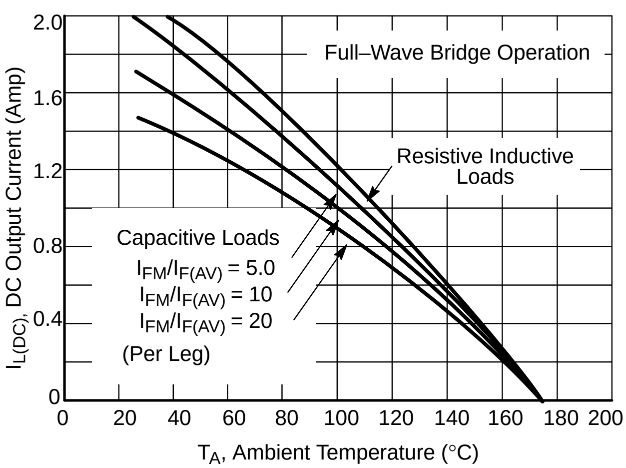


Figure 116. Derating Data for the Example

The diode bridge assembly has a surge current rating of 60 A for one cycle; i.e., it will handle two 60 A, half–cycle (8.3 ms) surges. The worst–case peak surge current in this application is

$$V_M/R_S = 50/1.0 = 50\text{ A}$$

The time of the surge is the time it will take to charge the capacitor, roughly the time constant of the series resistor and the filter capacitance, or

$$t \approx R_SC = (1.0)\,(500)\,\mu F = 0.5\text{ ms.}$$

Since the surge current is less than the rating and the time is much less than one cycle, the bridge will be satisfactory.

References

1. 1. J. Schaefer, *Rectifier Circuits,* John Wiley and Sons, Inc., New York, 1965, p. *258.*
2. 2. 0. H. Schade, *Analysis of Rectifier Operation,* Proc. IRE, Vol. 31, No. 7, July, 1943, pp. 343–344, 346–347.
3. 3. Robert H. Lee, *Origin and Characteristics of Harmonic Currents, Part I, Power Quality,* Volume 1, No. 5, 1990, pp. 349–351.
4. 4. Arthur W. Kelley and William F. Yadusky, *Rectifier Design for Minimum Line Current Harmonics and Maximum Power Factor,* Conference Proceedings of the Fourth Annual Power Electronics Conference and Exposition, Baltimore, 1989, IEEE Document number: 89Ch2719–3.

EQUATION 7. 1

is derived as follows:

$$\text{DC component} = \frac{V_M}{\pi}\int_{\omega t=0}^{\omega t=\pi} \sin\omega t\; d(\omega t) = \frac{2V_M}{\pi}$$

Ripple component of frequency $n\omega/2\pi$

$$= \frac{2V_M}{\pi}\int_{\omega t=0}^{\omega t=\pi} \cos\, n\omega t\;\sin\,\omega t\; d(\omega t)$$

$$= \frac{2V_M}{\pi}\left[\frac{\cos\,(n-1)\omega t}{2\,(n-1)} - \frac{\cos\,(n+1)\,\omega t}{2\,(n+1)}\right]_{\omega t = 0}^{\omega t = \pi}$$

$$= \frac{2V_M}{\pi}\left(\frac{-2}{n^2-1}\right)$$

In these equations *n* may have values 2, 4, etc.

Chapter 8

Rectifier Voltage Multiplier Circuits

Rectifier Voltage Multiplier Circuits

Voltage–multiplying power supply circuits are often employed when high–voltage sources are required, current demand is small, and regulation from the rectifier system is not too important. Voltage doublers are widely used in “universal input” line–operated equipment.

The principle of operation for all multiplying circuits is essentially the same. Capacitors are charged and discharged on alternate half–cycles of the ac supply voltage. The voltage at the output terminals is the sum of these voltages in series. Thus, the rectifier loading is necessarily capacitive, and high peak currents flow through the rectifier, much like rectifier circuits with capacitor input filters.

Voltage–Doubling Circuits

The conventional full–wave voltage doubler is shown in Figure 117. It is the most commonly used voltage–doubling, circuit. The circuit operates as follows: C_2 is charged to V_M through D_2 and R_{S2} during the negative half–cycle of the supply voltage. During the positive half–cycle, the supply voltage charges C_1 trough D_1 and R_{S1}. A voltage of 2 V_M appears across the output terminals. The capacitors’ voltage ratings must be greater than V_M, and the peak voltage rating of the rectifiers, V_{RRM} must be greater than 2 V_M. The ripple frequency is twice that of the input supply.

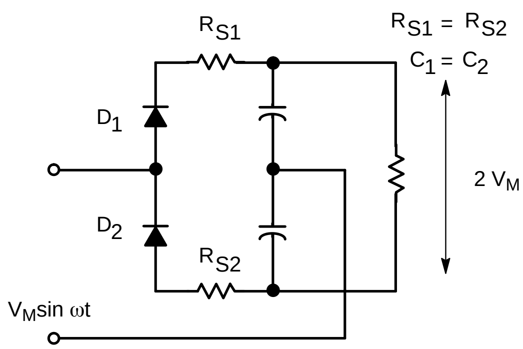


Figure 117. Conventional Full–Wave Doubling Circuit

Conventional voltage–doubler circuits may be designed from curves similar to those used for capacitor input filters. (See Figures 118 through 121.) Voltage doublers exhibit the high inrush and peak repetitive currents encountered with capacitive input filters. The source resistance, R_S, must be chosen to limit the inrush current to a safe level.

As an example, suppose approximately 200 V dc is required from the ac line to supply a l kΩ load. A rectifier is selected whose surge rating is 10 A. Ripple is not critical and it is desired to minimize capacitor size.

R_S is selected from:

$$R_S \geq V_{M(max)}/I_{FSM}$$
$$\geq (1.41)(130)/10 \text{ (High line is taken as 130 V)}$$
$$\geq 19\ \Omega \text{ (use 20 } \Omega\text{).}$$

Capacitor size is determined by using Figure 118, which requires that $V_{L(DC)}/V_M$ and R_S/R_L be calculated.

$$V_{L(DC)}/V_M = 200/(1.41)(117) = 1.21$$
(Normal line is taken as 117 V)
$$R_S/R_L = 20/1000 = 2\%$$
$$\text{Read } \omega CR_L = 7.0,$$
$$\therefore C = 7.0/\omega R_L = 7.0/(377)(10^3)$$
$$= 18.6\ \mu F \text{ (Use 20 } \mu F\text{).}$$

With C = 20 μF ωCRL = 7.54 and the remaining circuit information may be found:

From Figure 119, $I_{FM}/I_{F(AV)} = 5.5$

From Figure 120, $I_f/I_{F(AV)} = 2.2$

From Figure 121, r_f = 13%.

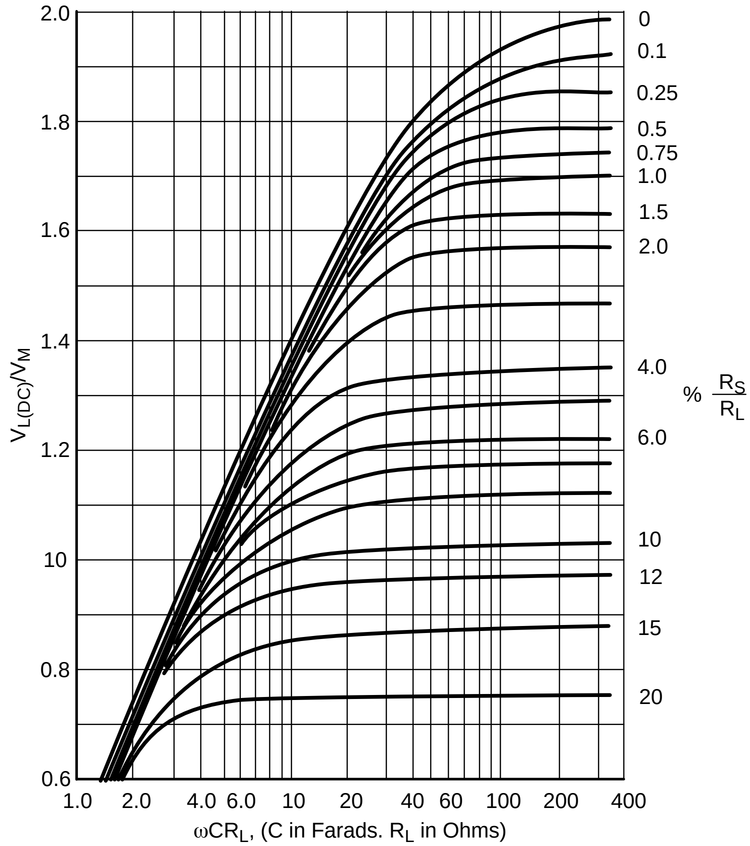


Figure 118. Output Voltage as a Function of Filter Constraints for Full–Wave Voltage Doubler

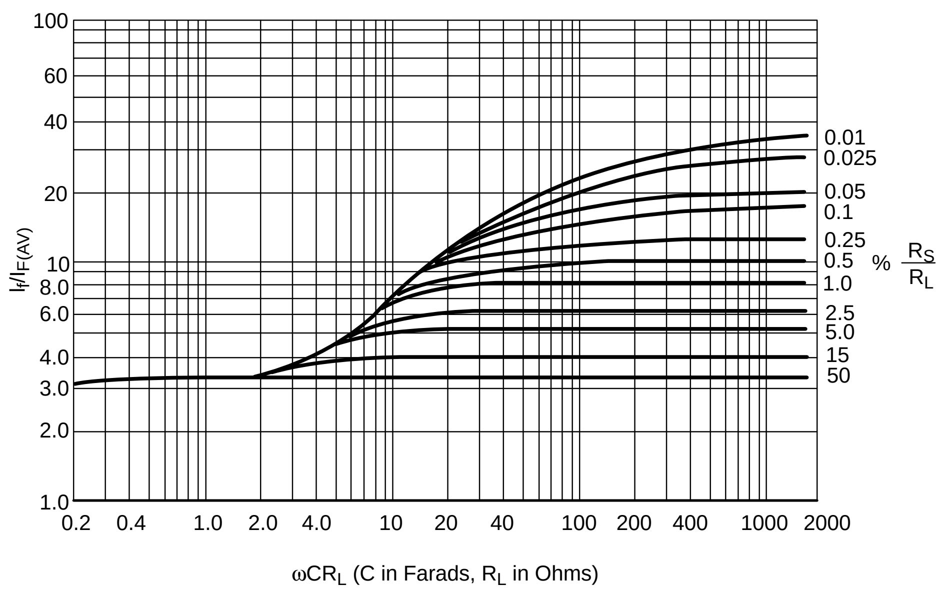


Figure 119. Peak Rectifier Current as a Function of Filter Constants for Full–Wave Voltage Doubler

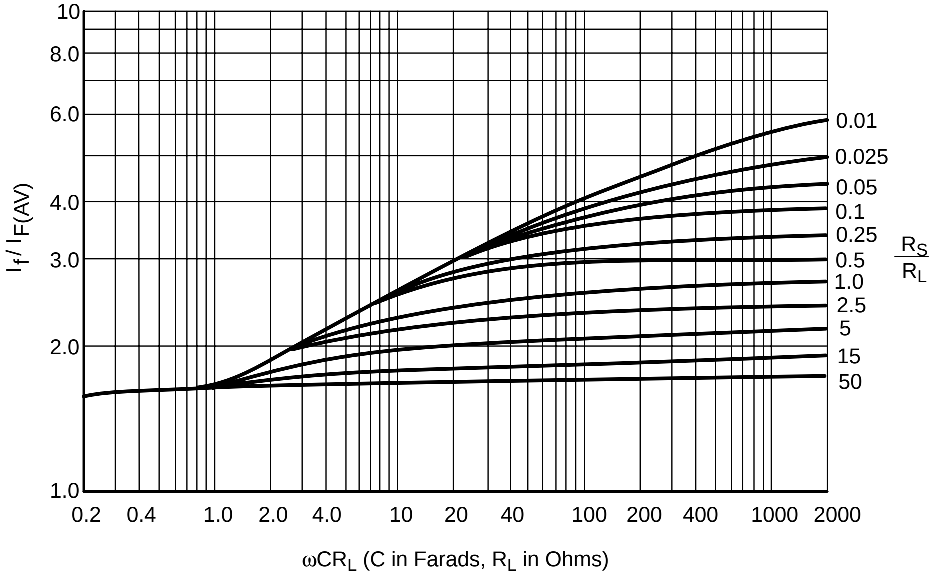


Figure 120. rms Rectifier Current as a Function of Filter Constants for Full–Wave Voltage Doubler

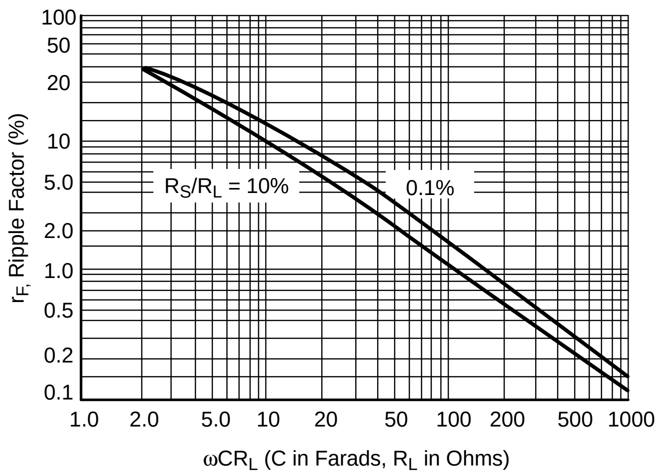


Figure 121. Ripple as a Function of Filter Constants for Full–Wave Voltage Doubler

In the cascade circuit (Figure 122), C_1 is charged through D_2 during the negative half–cycle. In the positive half–cycle, C_1 in series with the input voltage charges C_2 to a voltage of 2 V_M, when R_L is very large. The capacitor C_1 should be rated greater than V_M, C_2 greater than 2 V_M, and the rectifier should have a peak inverse voltage rating exceeding 2V_M. The ripple frequency is that of the input.

The cascade voltage doubler has poorer voltage regulation because the charge on C_1 must supply the charge to C_2; however, it has the advantage of a common connection between the output and the input. To compare the two voltage–doubling circuits, the circuits may be evaluated using similar components. The resulting characteristics are given in Table 22. Design curves have not been worked out for the cascade circuit. It performs better when C_1 is much larger than C_2.

Table 22. Voltage Doubling Circuit Characteristics

Circuit Configuration	V_{in} Volts (rms)	R_s Ohms	R_L Ohms	C μF	$V_{L(DC)}$ Volts	$I_{L(DC)}$ mA	r_f %
Conventional	117	20	1k	20	255	205	12.0
Cascade	117	20	1k	20	160	160	16.0

"Universal" Input Circuit

Switchmode power supplies normally rectify the ac power line mains directly without using a low–frequency transformer. Isolation from the line is provided by the high–frequency transformer in the power supply. Most manufacturers supply to the worldwide market place whose ac mains voltages and tolerances fall into two broad ranges: 90 to 130 V ac, and 180 to 270 V ac.

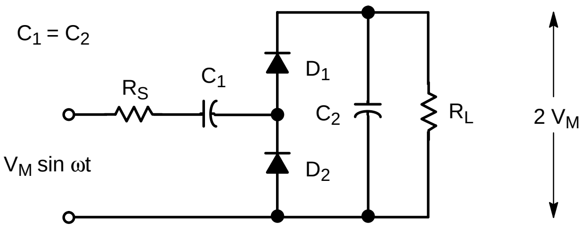


Figure 122. Cascade (Half–Wave) Voltage Doubling Circuit Full–Wave Voltage Tripler Circuit

A popular means of accommodating both ranges is shown by the circuit of Figure 123. When the switch is open, the circuit operates as a full bridge rectifier. With 230 V ac input, the dc output is approximately 320 V. When the switch is closed, the circuit is configured as a full–wave voltage doubler, producing about the same output with a 115 V ac input. The switching power supply is designed to provide a regulated output despite variations in the dc level caused by operation over either ac input range.

Switching can be made automatic [1] in response to the applied line voltage in order to eliminate power supply failure caused by having the switch or terminal straps in the wrong position. For automatic switching, a triac is used for the switch. It is turned on by a circuit which senses when the ac line peaks are below that produced by 130 V rms. To simplify the design, integrated circuits such as the MC 3423P overvoltage sensing circuit are available.

The ripple is highest when the circuit is operating in the voltage–doubling mode. To conserve space and lower cost, small capacitors are desirable. Therefore, high ripple on the order of 10 to 20% is often allowed; the switching supply reduces the ripple to a low value in its outputs. On the other hand, in high–quality equipment designed to ride through a half or full cycle line voltage dropout, the filter capacitors may be somewhat oversized in order to provide the desired hold–up time.

The preceding example illustrated the use of the curves when maximum output voltage was not a design goal but it was desired to minimize the capacitor size. In most cases, particularly for the universal input, ωCRL is chosen on the order of 50 or more and R_S/R_L is as small as practical in order to obtain a high output voltage. For example, with $\omega_{CRL} = 100$ and $R_S/R_L = 1\%$, $V_{L(DC)}/V_M$ is only 1.7 (from Figure 118). To more nearly approach 2, R_S/R_L ratios are chosen on the order 0.25%. With $R_S/R_L = 0.25\%$ and $\omega_{CRL} = 100$, $V_{L(DC)}/V_M$ is a much more acceptable ratio of 1.85. However, the rectifier peak–to–avenge current ratio is now about 11. (See Figure 119) When switched to full bridge operation, the filter capacitors are placed in series dropping ω_{CRL} to 50. The curves in Chapter 7 reveal that $V_{L(DC)}/V_M = 0.96$ and $I_{FM}/I_{F(AV)} = 15$, worse than during operation as a doubler.

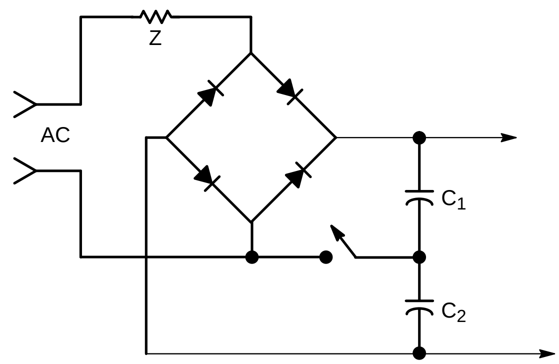


Figure 123. Dual Input Voltage Range Circuit

During turn–on, the low value of R_S needed for good doubler operation results in enormously high inrush current. Consequently, universal rectifier circuits must use some method of having R_S high at turn–on and low during steady–state operation. Two schemes are presently used; one uses a negative temperature coefficient thermistor and the other uses a triac.

The thermistor is used in place of R_S on Figure 123 and is chosen to have a room temperature resistance to limit inrush current to a safe value. As charging current flows, the thermistor heats; its temperature rise causes a corresponding drop in its resistance value. The time constant of the thermistor must be longer than that of the rectifier filter for this scheme to work. The disadvantage of the thermistor is that it cannot respond to momentary interruptions on the ac line which allows a high surge current to flow when the line voltage reappears.

A more foolproof and commonly used scheme parallels R_S – which is large – with a triac. A control circuit fires the triac when certain circuit conditions are met which insures that the capacitors are fully charged.

Voltage Tripling Circuits

The full–wave voltage tripler shown in Figure 124 operates as follows. On the negative half–cycle of the supply voltage, C_1 and C_3 (in parallel) charge to V_M. On the positive half–cycle C_1 (in series with the supply voltage) charges to $2V_M$. Thus a voltage of $3V_M$ will appear at the output terminals of the circuit. The voltage rating of C_1 and C_3 must be greater than V_M, and the rating of C_2 must be greater than $2V_M$. The diodes should be able to block at least twice the maximum peak value of the input supply voltage. The ripple frequency of the output is twice that of the input. The output characteristics for the full–wave tripler are shown in Figure 125.

The cascade principle may be extended to obtain a voltage tripler as shown in Figure 126. The operation of the cascade voltage tripler is quite similar to that of the cascade voltage doubler circuit. The voltage rating of C_1 should be greater than V_M and the voltage ratings of C_2 and C_3 should be greater than $2V_M$. The peak inverse voltage ratings of the diodes should also exceed $2V_M$. The ripple frequency of the output voltage will be the same as the frequency of the applied voltage.

As might be anticipated, the regulation and ripple of the cascade voltage tripler is inferior to that of the full–wave voltage tripler. For a comparison of the two circuits, refer to Table 23.

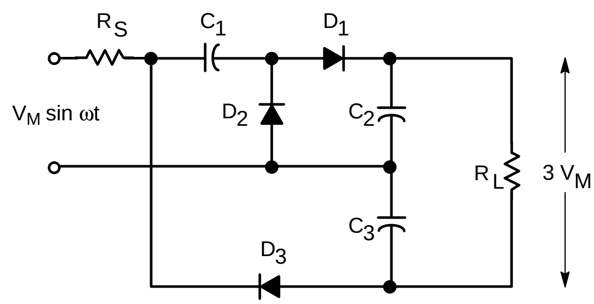


Figure 124. Full–Wave Voltage Tripler Circuit

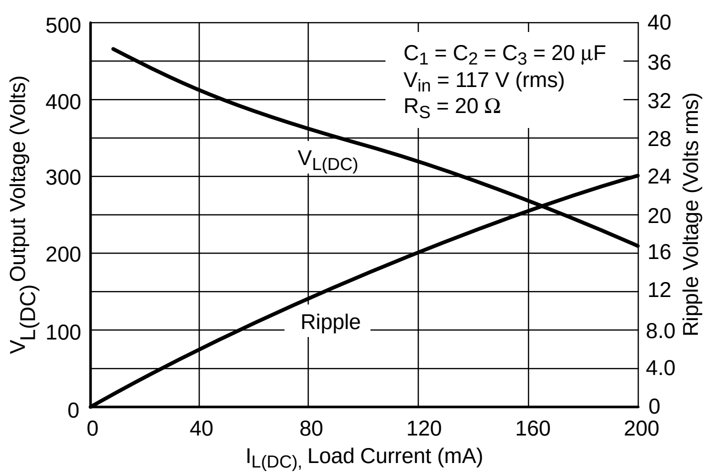


Figure 125. Output Characteristics for the Full–Wave Voltage Tripler Circuit of Figure 124.

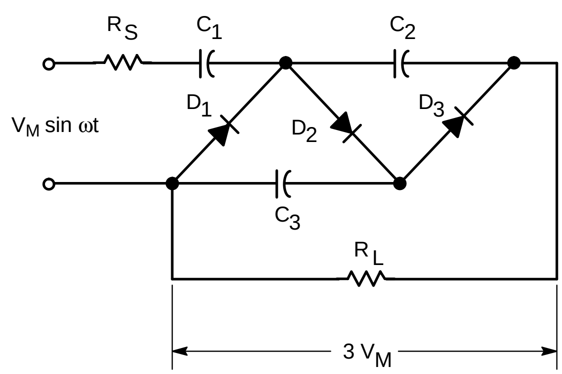


Figure 126. Cascade (Half–Wave) Voltage Tripler

Table 23. Voltage Tripling Circuit Characteristics

Circuit Configuration	RL Ohms	$V_{L(DC)}$ Volts	$I_{L(DC)}$ mA	r_f	RL Ohms	$V_{L(DC)}$ Volts	$I_{L(DC)}$ mA	r_f %
Full–Wave	5.0 k	385	77	2.4	1.5 k	255	170	8.0
Half–Wave	5.0 k	328	66	28.0	1.5 k	170	112	37.0

Voltage Quadruples

The full–wave voltage quadrupler circuit of Figure 127 operates in the following way: Assume that the input supply voltage is positive. This causes capacitor C_1 to charge through D_2 to V_M. Also, capacitors C_4 and C_2 in series with D_4 charge to $V_M/2$ (assuming the capacitors are equal in value). The charging voltage polarities of C_1, C_2 and C_4 at this point are shown in Figure 127. On the negative half–cycle, C_1 in series with the supply voltage charges C_3 through D_1 to 2 V_M. Also, capacitor C_2 charges to V_M through D_3 with the opposite polarity than that shown in Figure 127. On the next positive half–cycle, capacitor C_4 is charged to 2 V_M since the voltage across C_2 and the supply voltage are in series. In addition, the charge on C_1 is replenished. At this point, the output voltage across C_3 and C_4 is $4V_M$. The next negative cycle replenishes the charge on C_2 and C_3 and the next positive cycle replenishes the charge on C_1 and C_4.

The ripple on the output has a frequency twice that of the input. The voltage rating of C_1 and C_2 must be greater than V_M while that of C_3 and C_4 must be greater than 2 V_M. All of the diodes must have a voltage rating greater than 2 V_M. Output characteristics are shown in Figure 128.

The cascade half–wave voltage quadrupling circuit is shown in Figure 129. The operation of this circuit is basically the same as the other cascade networks. The capacitance of C_1 should be high and its voltage rated above V_M while the remaining capacitors and the diodes should be rated above 2 V_M. The ripple frequency of the output is the same as that of the supply.

The regulation of the half–wave voltage quadrupler is inferior to the full–wave as might be expected. (Compare Figure 128 to Figure 130.) To improve regulation, the values of capacitors C_1 and C_2 are increased. To illustrate this improvement, the output characteristics of the circuit of Figure 129 were obtained when all capacitors equal 20 µF and 40 µF The results are shown in Figure 130.

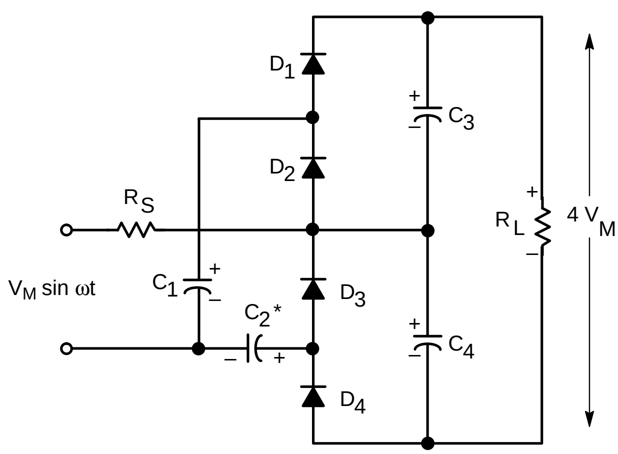


Figure 127. Full–Wave Voltage Quadrupler
*Voltage Polarity is opposite to that shown when fully charged. See text.

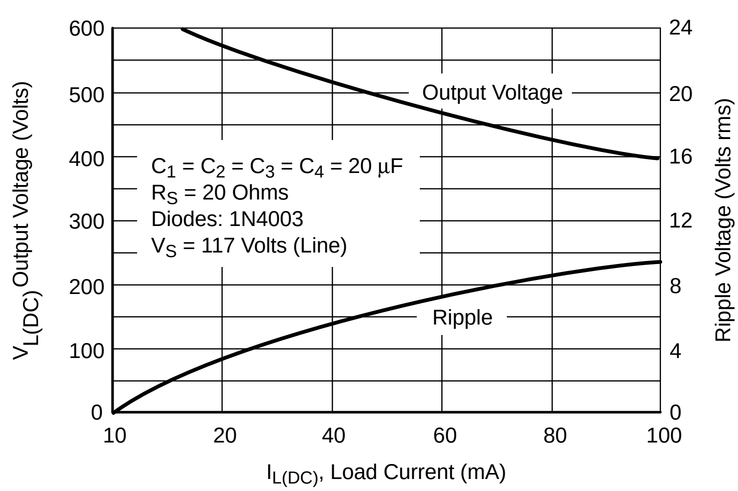


Figure 128. Output Characteristics of a Full–Wave Voltage Quadrupler Circuit.

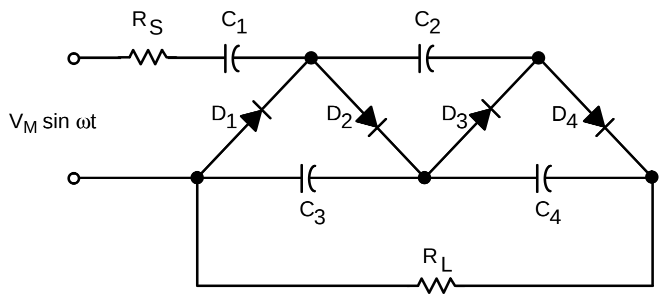


Figure 129. Cascade (Half–Wave) Voltage Quadrupler Circuit

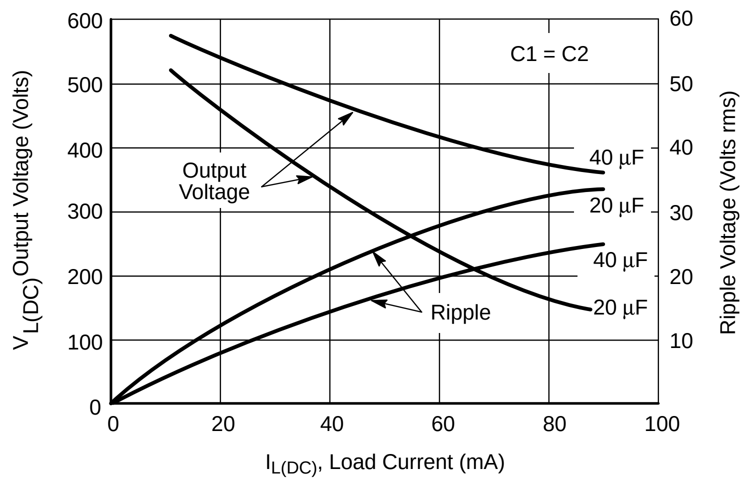


Figure 130. Output Characteristics of the Half–Wave Voltage Quadrupler Circuits

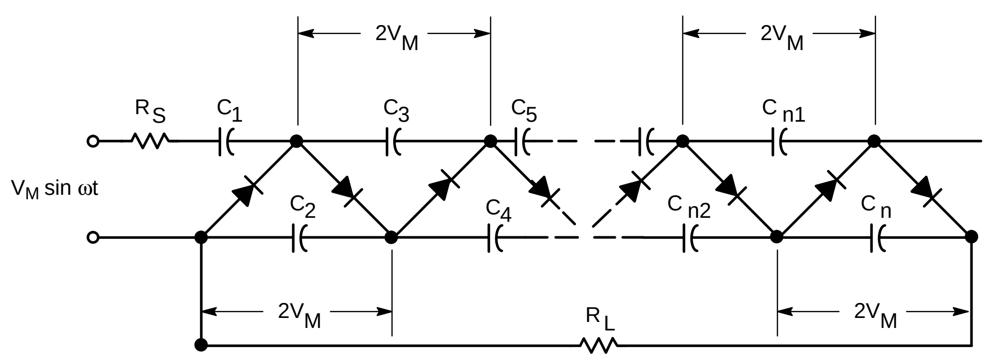


Figure 131. n–Section Voltage Multiplier Circuit

High–Order Cascade Voltage Multipliers

In theory, the number of sections which can be added to the basic cascade circuit is unlimited, as illustrated in Figure 131. In practice, the number of sections is limited because the regulation becomes progressively worse as additional sections are added.

Both even and odd multiples of VM can be obtained. The capacitor connected to the supply charges to VM. The remaining capacitors on that side of the circuit charge to 2 VM. Thus, odd multiples of VM from the upper branch and even multiples from the lower branch can be obtained with respect to ground.

References

1. S.K. Tong and K.T. Cheng, *External–Sync Power Supply with Universal Input Voltage Range for Monitors*, ON Semiconductor Application Note, AN1080.

Chapter 9

Transient Protection of Rectifier Diodes

Transient Protection of Rectifier Diodes

In order to achieve the high degree of reliability already associated with silicon rectifier diodes, it is imperative to understand that the ratings are based on the absolute maximum rating system. With this system, ratings apply within a certain defined set of operating, test, or environmental conditions, and degradation or failure may be caused by a combination of electrical or environmental stresses, even though application of any one rating will not impair the performance of the device. Ratings are established by extensive tests, including life–tests conducted by the manufacturer, and are typically checked during production by quality control sampling. The user should be aware that there is no safety factor implied above each maximum rating, although manufacturers generally employ one.

Because of the intolerance of silicon rectifier diodes to voltage transients in excess of their ratings, a proper circuit design generally requires some built–in means to afford safe operation. Therefore, it is imperative to discuss some of the causes of voltage transients and means of limiting them. Unwanted transient voltages, commonly referred to as surges, are present in all electrical and electronic systems. The origin of surges may be external to the electronic equipment or surges may be internally generated by the equipment's electronic circuits. Surges are randomly created on the utility distribution system by lightning, capacitor bank switching, start–up and shut down of electrical and electronic equipment, and fault clearing. These surges often have high energy and must be considered in the design of rectifier circuits operating from the utility mains. Internally generated surge voltages are usually caused by rapid switching of current. The surge voltage produced is a result of energy stored in circuit and wiring inductance.

Overcurrent can also lead to failure of semiconductor diodes. The small thermal mass associated with silicon devices allows the junction temperature to rise and fall very rapidly during a current surge and remain undetected by an external thermocouple. If, during this event, the junction temperature exceeds the maximum rated value, the device may fail destructively (short) as a result of thermal breakdown.

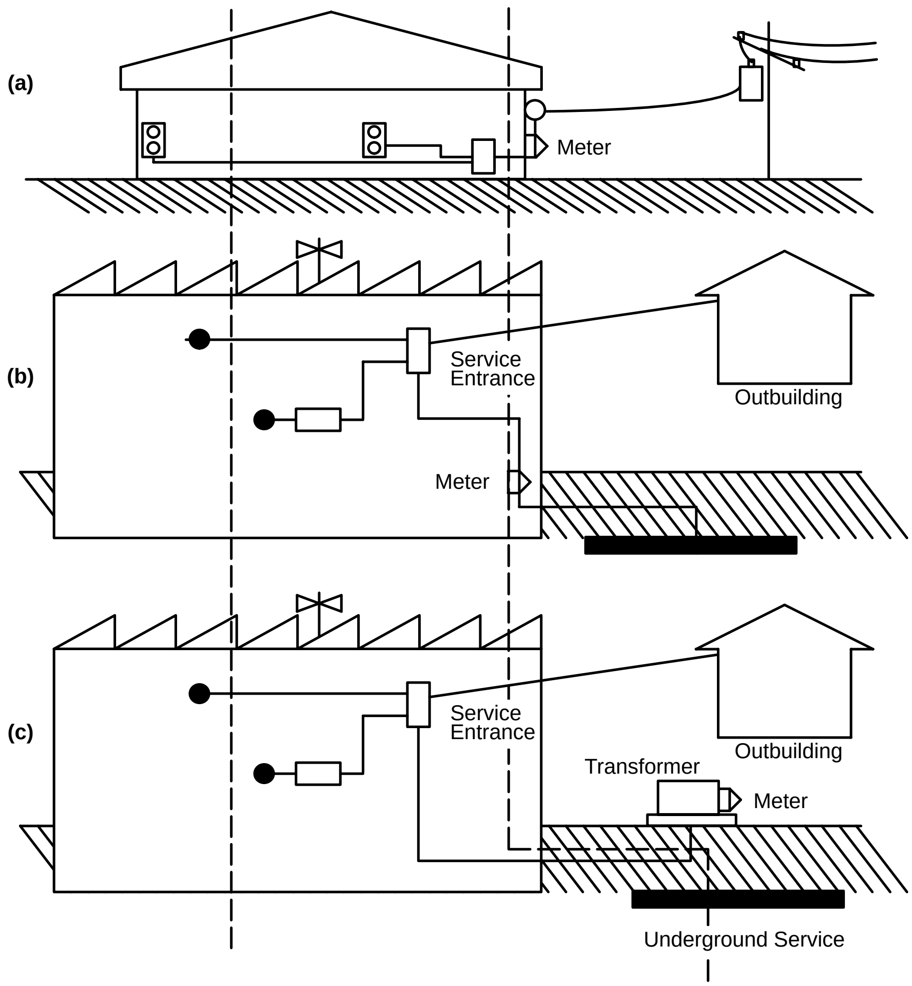
Consequently, silicon diodes are specified so that the excursion of the junction temperature during expected overloads can be calculated and thus held within proper limits through circuit design and thermal management. Fuses, circuit breakers, or protective circuitry are used to protect rectifier diodes from excessive current.

Externally Generated Voltage Surges

Several years of work by many people in IEEE committees culminated in the publication of IEEE standard 587–1980, which describes "typical" transient conditions occurring in unprotected power circuits. It addresses transient voltages that exceed twice the peak operating voltage, with durations ranging from a fraction of a microsecond to a millisecond, and originating primarily from system switching and lightning effects. Transients shorter than 100 ns and long–duration surges that sometimes exist on power lines are not covered by the standard. The standard also proposes tests which approximate the real–world transient conditions for the purpose of evaluating the survival capability of equipment connected to ac mains. The IEEE standard was later adopted by ANSI and given the document number C62.41.

A major revision of the standard was issued in 1991, upgrading it to a "recommended practice." [1] Designers of power–line operated equipment should familiarize themselves with its contents. The revised document, while retaining most of the original information, provides updated and expanded information relevant to "typical" surge environments based upon location within a building. Additional guidance is given on suggested specifications to be used for equipment, based upon exposure level. The material which follows is based upon the main points of the recommended practice, applicable to the design of a line–operated rectifier system.

Three location categories are defined: "A" and "B" for indoor applications, and "C" for outdoor applications. The location categories are further defined in Figure 132. They take into consideration the increase in power–line impedance to high–frequency surges from the outside to locations well within the building. Most consumer electronic equipment can be expected to find use in location category A while industrial commercial equipment is often likely to be used in category B locations.



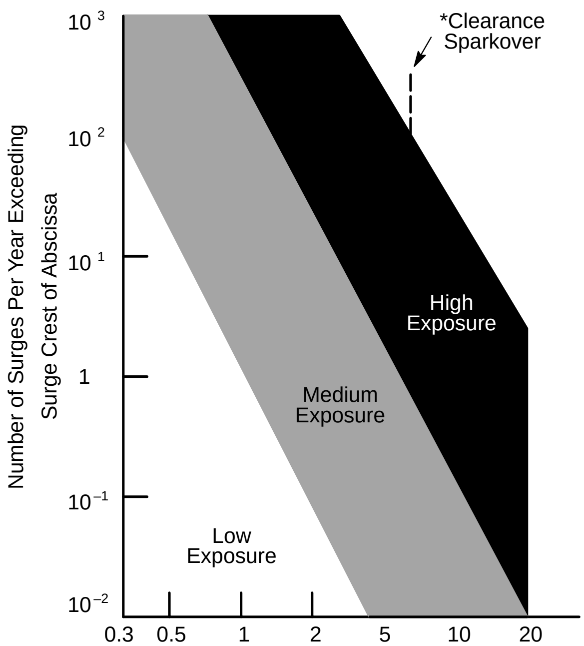
A	B	C
Outlets and long branch circuits.	Feeders and short branch circuits	Outside and service entrance
All outlets at more than 10 m (30 ft) from Category E	Distribution panel devices	Service drop from pole to building
All outlets and more than 20 m (60 ft) from Category C	Bus and feeder in industrial plants	Run between meter and panel
	Heavy appliance outlets with “short” connections to service entrance	Overhead line to detached building
	Lighting systems in large buildings	Underground line to well pump

Demarcation between Location Categories B and C is arbitrarily taken to be at the meter or at the mains disconnect (NEC Art. 230–70 [B2]) for low–voltage service, or at the secondary of the service transformer if the service is provided to the user at a higher voltage.

Figure 132. Location Categories

A summary of data taken on the rate and magnitude of transient voltage occurrence is provided. Transient occurrence varies over wide limits depending upon the power system, its loading, and the amount of lightning activity. Data collected from many sources have led to the plot shown in Figure 133. This prediction shows only the relative frequency of occurrence, while the number of occurrences can be described only in terms of low, medium, or high exposure. These exposure levels are defined in general terms as follows:

1. Low Exposure. Systems with little load switching activity, that are in geographical areas known for low lightning activity.
2. Medium Exposure. Systems in geographical areas known for medium–to–high lightning activity and/or significant switching transients.
3. High Exposure. Rare but real systems subject to severe switching transients and/or extensive lightning activity.



*In some locations sparkover of clearances may limit the overvoltages to lower levels

Figure 133. Annual Rate of Occurrences at Unprotected Locations

Two recommended standard test waveforms are defined as shown in Figure 134. The first test wave is called a combination impulse wave; its open circuit voltage and short circuit current are shown in Figure 134(a) and Figure 134(b) respectively. The second test wave is a ring wave as shown in Figure 134(c).

Impulse waves have been in use for many years to test lightning arrestors and high–power surge suppression devices, because impulses are characteristic of the waveform produced by lightning strikes to power lines. The short form nomenclature for an impulse wave is t_r/t_d where t_r denotes the rise time and td denotes the pulse width, defined at the point where the amplitude has decayed to 50% of its peak value. Thus the combination wave has a 1.2/50 μs open–circuit voltage and an 8/20 μs short–circuit current. The impulse wave deposits substantial energy in a surge–protective device and is usually chosen to specify protector performance. For example, Underwriters Laboratories (UL) has chosen the standard's combination wave to test portable power taps (i.e., power strips) having surge–protective devices.

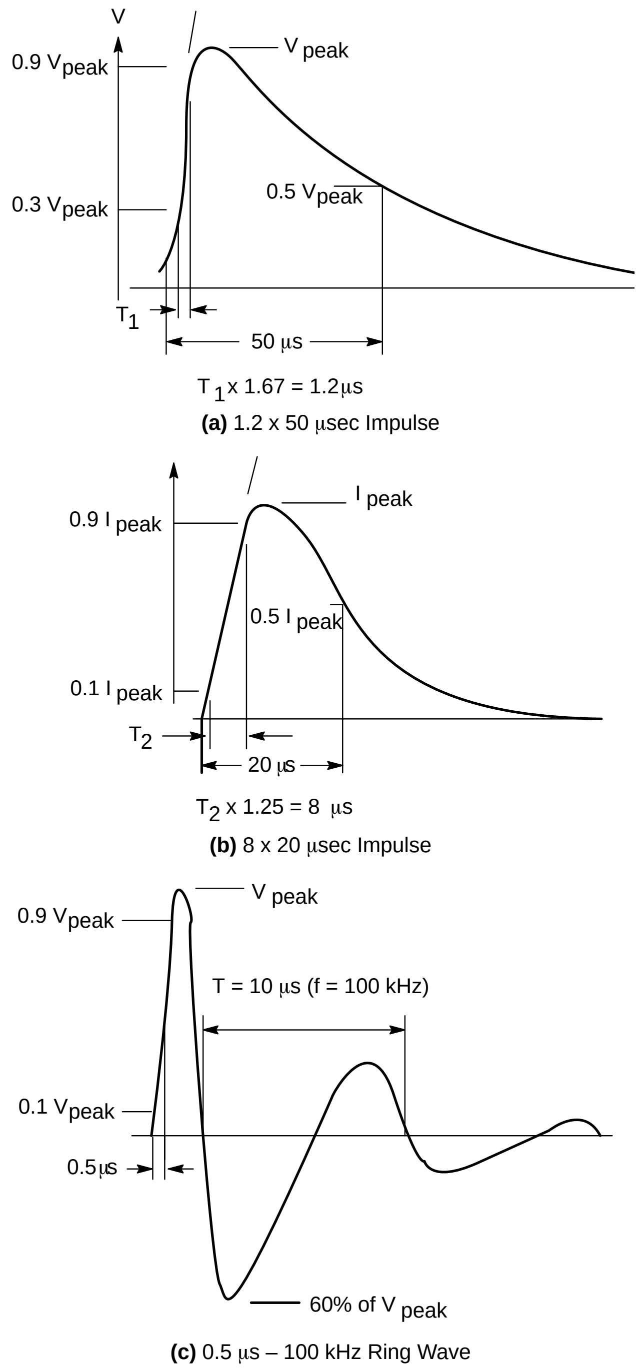


Figure 134. Standard Test Waveforms Defined

Based upon a consensus of engineering experts, the amplitudes expected for various combinations of location categories and exposure levels are shown in Table 24. The information is taken from the updated version of C62.41; in the original version, only the high exposure levels were suggested. The revision recognizes the significance of system exposure as well as equipment location categories.

The ring wave is typical of most disturbances within a building because transient energy excites the natural resonant frequency of the wiring system. In practice, the frequency can vary from a few kilohertz to a few hundred kilohertz. The fast rise time is used because equipment using thyristors is prone to false triggering with high d_v/d_t, and fast rising wavefronts can cause a nonuniform voltage distribution in windings. The 0.5 μs/l00 kHz ring wave is therefore primarily aimed at evaluating equipment immunity to fast wavefronts and high–frequency disturbances. Since the ring wave is not intended to check energy capability, its effective source impedance is chosen much higher than that of the combination wave.

Fast–rising surges are produced on power lines by opening airgap switches such as relays and contactors. Although fast transients are attenuated rapidly with distance by the impedance of the power line, equipment in close proximity may be affected. A common situation occurs when two pieces of equipment are powered from the same duplex receptacle. Switching off the power to one will couple a high–frequency transient to the other, possibly causing a malfunction. The waveform chosen to represent the fast transient is a 5/50 ns impulse applied in a burst having a duration of 15 ms. It is delivered by a generator having a 50 Ω source impedance with a peak amplitude of 1 kV, 2 kV, or 4 kV depending upon the severity level desired.

High–energy surges have been added to the menu of waveforms. The 10/1000 μs impulse is suggested as typical of lightning–induced pulses traveling in long underground cables, surges generated by fuse blowing in long cables, and the envelope of the oscillation caused by capacitor bank switching. A 5 kHz ring wave is also suggested as typical for the oscillation caused by capacitor bank switching.

The levels and impedance values chosen for testing with the high–energy surges are shown in Table 25. Building location categories do not apply because the building wiring does not significantly attenuate the low–frequency surges. Levels of test voltages and source impedance of the surges are based upon system exposure and the system voltage. For example, the peak total voltage developed by fuse clearing occurring near the peak of the power line voltage, for a 120 V rms line in an industrial environment, is the sum of peak line voltage and peak transient voltage. That is:

$$V_{total} = 120\sqrt{2} + (1.3)(120)\sqrt{2} = 390\ V$$

In reviewing the data presented, it is obvious that the transient voltage levels are highly site dependent. The dilemma a designer faces is to provide sufficient transient protection to equipment to insure a reasonable survival rate without adding excessive cost, although the location where the equipment will be used and the severity of the exposure level is not known. A reasonable solution to the dilemma is to design for a category B, medium exposure level. Users located in high exposure sites have no doubt experienced troubles with all electronic equipment and have installed surge protective devices throughout the facility.

Table 24. Suggested Levels of Standard Voltage and Current Surges

Location Category	System Exposure	Peak Voltage	Ring Wave		Combination	
			Peak Current	Effective Impedance	Peak Current	Effective Impedance
A	Low	2 kV	67 A	300	–	–
	Med	4 kV	133 A	300	–	–
	High	6 kV	200 A	300	–	–
B	Low	2 kv	167 A	120	1 kA	Ω
	Med	4 kv	333 A	120	2 kA	Ω
	High	6 kV	500 A	120	3 kA	Ω
C	Low	6 kV	–	–	3 kA	Ω
	Med	10 kV	–	–	5 kA	Ω
	High	20 kV	–	–	l0 kA	Ω

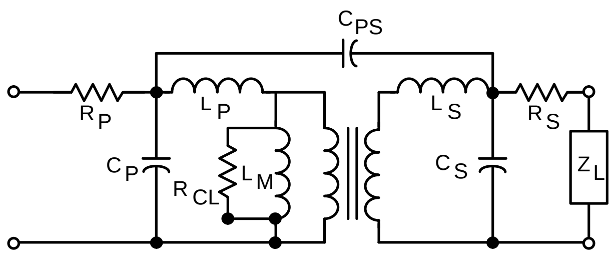
Table 25. Suggested Levels for High Energy Surges

System Exposure	10/1000 μs Impulse			1.5 μs/5 kHz Ring Wave		
	Exposure Definition	Peak Voltage	Surge Impedance	Exposure Definition	Peak Voltage	Surge Impedance
Low	Residential	–	–	Far from Switched Bank	–	–
Medium	Commercial	1.0 V_{pk}	1 Ω	Typical	1.0 V_{pk}	1–5 Ω
High	Industrial	1.3 V_{pk}	0.25 Ω	Near Large Switched Banks	1.8 V_{pk}	0.5–1.0 Ω

Internally Generated Voltage Surges

Normal equipment operation generates surge voltages. In practice, the possibilities are almost endless; the circuit designer must astutely analyze the transient–generating potential of a particular system in their electrical environment. A few situations, however, are quite common.

When transformers feed rectifiers from the ac power line, several mechanisms for generating transients exist. A generalized lumped circuit model for a transformer is shown in Figure 135. It should not be taken too literally because the elements shown are actually distributed in nature. However, the model allows several effects to be visualized.



R_{CL} = core loss representation
L_m' : primary magnetizing inductance
R_p : primary winding resistance
L_p : primary winding leakage inductance
T : ideal transformer
L_s : secondary winding leakage inductance
R : secondary winding capacitance
ZL' : secondary load impedance
C_{PS} : primary to secondary interwinding capacitance
C_P : primary winding capacitance
C_S : secondary winding capacitance

Figure 135. Broadband Circuit Model of a Transformer

Note that the interwinding capacitance (C_{PS}) and the secondary winding capacitance (C_S) form a capacitive divider. Depending upon transformer construction, their ratio may vary widely. The winding techniques commonly used for 60 Hz applications allow C_{PS} to exceed C_S. Consequently, if a voltage step is applied at the input, such as with a switch closure occurring at the peak of the ac line, it is possible for a voltage surge almost equal in amplitude to the peak ac primary voltage to appear across the secondary terminals. Since the surge is not directly related to the turns ratio, the rectifier reverse voltage may be many times the normal peak inverse voltage when a stepdown transformer is used. It should be evident that any line voltage surge with a fast rise time will also be coupled through the interwinding capacitance and appear at the secondary without much attenuation.

In addition, because of system resonance the input voltage step usually generates a damped oscillatory transient with a peak amplitude possibly as high as twice the normal secondary voltage. The frequencies of oscillation are determined by the values of the parasitic reactances in the transformer and its load. The effects just described are illustrated in Figure 136.

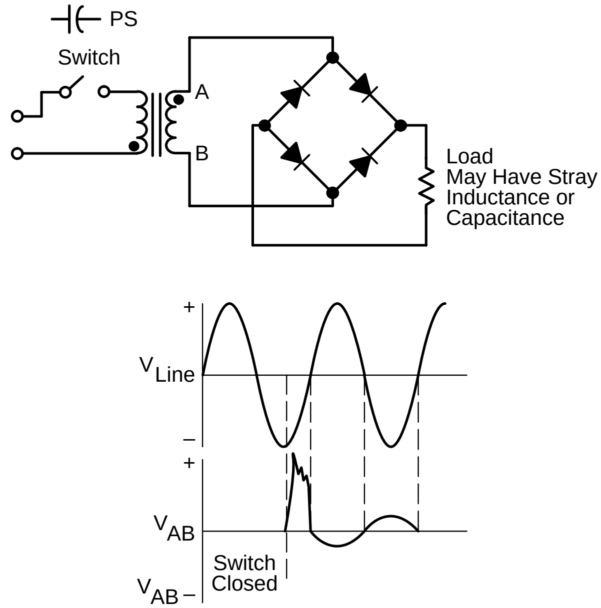


Figure 136. Situation where Transformer Interwinding Capacitance Couples a Transient to the Secondary

Opening the primary circuit may generate even more extreme voltage transients resulting from the collapse of the magnetic flux in the core, represented by L_M in Figure 135. A worst–case situation is shown in Figure 137 where switch opening occurs at the time of zero voltage crossing of the ac line. At this time, the magnetizing flux is at its maximum and its collapse may produce a transient, an order of magnitude higher than the normal secondary voltage. In practice, the generated surge voltage is not as clean as shown because of switch arcing.

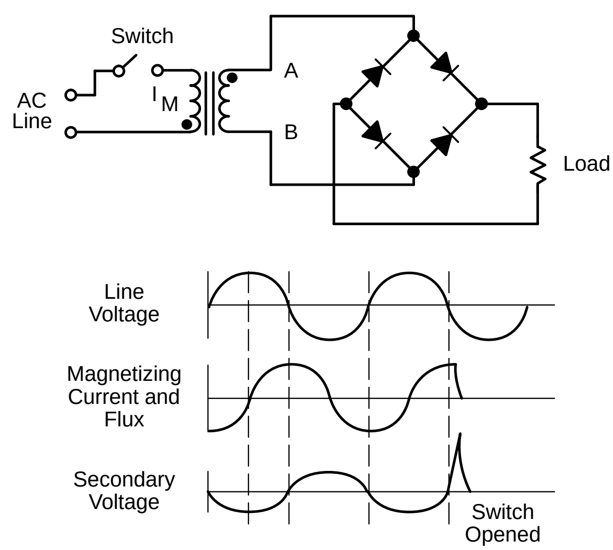


Figure 137. Typical Situation Showing Possible Transient when Interrupting Transformer Magnetizing Current

Switching of loads on the output of a rectifier can cause transients to be developed across the diodes, as shown in Figure 138. The inductance L_{EQ} represents the equivalent inductance of the power source, consisting of wiring inductance plus leakage inductance if a transformer is present. A short across Load 2 causes a very high current to be drawn which stores a large amount of energy in L_{EQ}. When the fuse blows causing a rapid decrease in current, the inductance generates a voltage of Ldidt having an energy of $LI^2/2$. A similar but less severe effect occurs when normal loads are intentionally switched off.

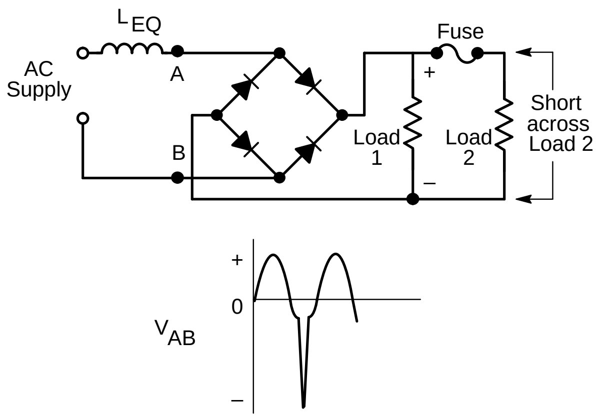


Figure 138. Transients Caused by Load Switching

Surge Protection Circuits

Effective control of transient voltage surges in equipment used on ac power lines invariably uses a metal oxide varistor (MOV). It can handle a large amount of transient energy for its weight, size, and cost. Service life is long, if properly applied. MOV characteristics are shown in Figure 139 for a part suitable for use across a 120 V ac power line. It is bi–directional and exhibits a fairly sharp breakdown, which helps limit voltage surges.

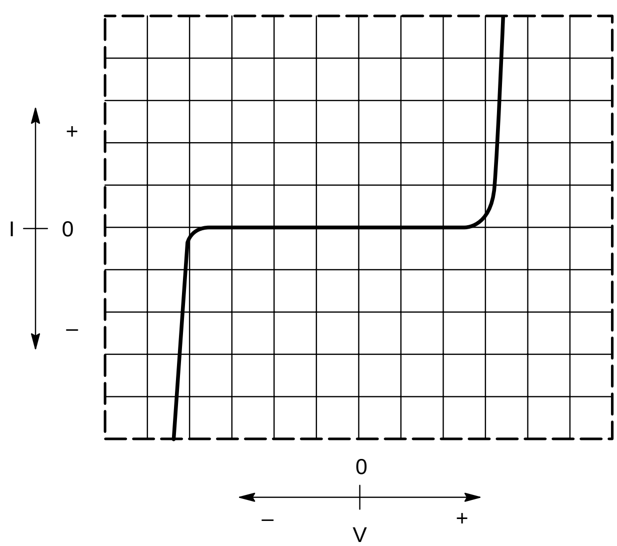


Figure 139. Typical V–I Characteristic of an MOV

Transients from lightning and many other sources enter the power system as a current source. The high surge voltages observed across power lines are the result of high surge currents flowing through the loads across the line. In order to prevent generation of high voltages, it is generally desirable to use an MOV directly across the line to divert the surge current back to its source, as shown in Figure 140. If the voltage across the varistor is too high for the downstream electronics, an inductor is used to restrict the current flowing into the rest of the circuit. A second MOV or other voltage–limiting device is used to hold the voltage at the circuit to the desired level. When a rectifier circuit with a capacitor input filter is in a medium exposure environment, it may perform the function of the second clipper.

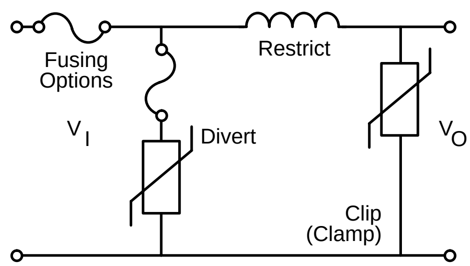


Figure 140. Functions of Elements Used in a Protection Network

MOVs are well suited to handle the l00 kHz ring wave and the 8/20 μs impulse associated with the majority of transient activity. But they cannot handle the high energy of the less frequent long waves typical of fuse blowing, capacitor bank switching, and line voltage swells. If the downstream circuits can be designed to handle these long duration surges, which may approach twice nominal line voltage, MOVs which break down at twice the line voltage peak can be used. This approach has a long history of success in the design of semiconductor based systems used in electric utility applications.

In cases where the downstream circuits cannot be designed to be cost–effective with a twofold voltage margin, MOVs which clamp at lower levels may be used; however, the MOV is likely to become a sacrificial element when a long–duration surge is encountered. The designer is faced with trading off the cost advantages of using lower voltage downstream components against the increasing probability of protective device failure as the clamping voltage is lowered. [2] The breakdown voltage of the varistor, as an absolute minimum, must be above the peaks of a high line voltage. For example, a l40 $V_{(rms)}$ rated MOV is the minimum suitable for a 120 V nominal power line in the U.S.

Fusing the equipment is important because when an MOV fails, it fails as a short. However, the resulting fault current will open the device, possibly violently.

Two fusing options are shown in Figure 140. Placing the fuse in series with the input line insures that the equipment is never operated without protection. If the line fuse is sized to handle high power equipment or equipment having a substantial inrush current, the varistor may rupture when shorted before the fuse has time to open. To avoid this situation, a rugged varistor should be selected which can handle the line fault current for enough time to cause the line fuse to blow.

Placing the fuse in series with the MOV permits the equipment to continue operating but without protection. This approach is risky, but has merit when the downstream circuits are able to handle some transient voltages without degradation. An additional risk introduced when the fuse is used in series with the varistor, is that of the surge caused by fuse blowing. During the time the varistor is shorted, sufficient energy may be stored in the upstream line inductance to cause a damaging voltage surge at the input terminals to the rectifier circuit. To minimize the surge at fuse blowing, a fast–acting or current–limiting fuse should be used.

An example of applying surge suppression to a three–phase circuit is shown in Figure 141. The MOVs across the input lines serve as the diverters. The filter choke provides the restrict function, and the filter capacitor also serves as the clipper. Since these filter components are large, the surge voltage limited by the MOVs is greatly attenuated at the dc output. Consequently, it is usually possible to select the MOV breakdown voltage to exceed twice that of the nominal line voltage peaks. The inductor must be effective for the 8/20 μs pulse. It will need to be wound to minimize feed–through capacitance, or else a small high–frequency inductor can be placed in series with the low–frequency filter inductor. The circuit should be tested to determine how much surge voltage is coupled through the transformer; the surge may not follow the turns ratio.

The popular “universal input” single–phase rectifier adapted to include transient protection is shown in Figure 142. Tests have shown the importance of matching the varistor voltage to the incoming line voltage and have indicated typical performance when a 2000 A, category B, 8/20 μs current surge is applied [3]. With R_1 = 0.5, L_1 = 100 μH, and $C_1 = C_2$ = 540 μF, the 2000 A surge caused the 320 V dc output voltage to rise by 74 V. In many cases, the following circuits can be designed to handle this extra voltage.

The inductor may be used as part of the EMI filter commonly required on equipment. It is also possible to use a second section of suppression to make the circuit almost immune to transients, while the inductors also serve as the normal mode and common mode EMI filters [4]. Such a circuit is shown in Figure 143. With a 3000 A 8/20 μs surge, the transient voltage on the output filter capacitors is a mere 10 V. The peak inverse voltage across the rectifier is determined by the peak limiting voltage of the MOV The EMI filter provided over 100 dB of attenuation from 6 kHz to 50 MHz, differential mode, and from 6 kHz to 12 MHz common mode. Good filter and surge suppression performance is dependent upon the coils. They need high Q over a broad frequency band and must be wound to minimize shunt capacitance.

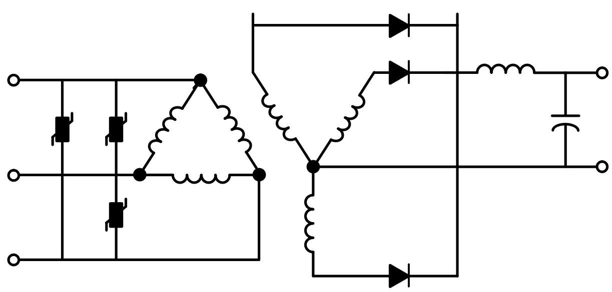


Figure 141. Transient Suppression in a Three–Phase Rectifier

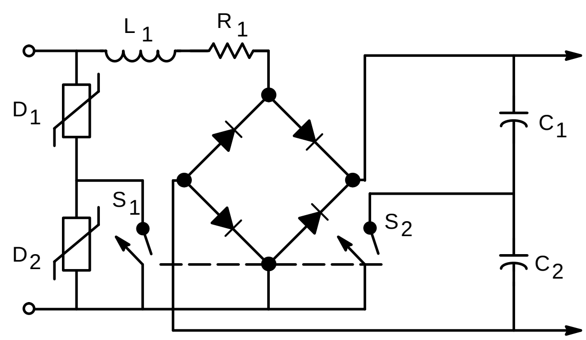


Figure 142. MOV Protection for a Universal Input Rectifier

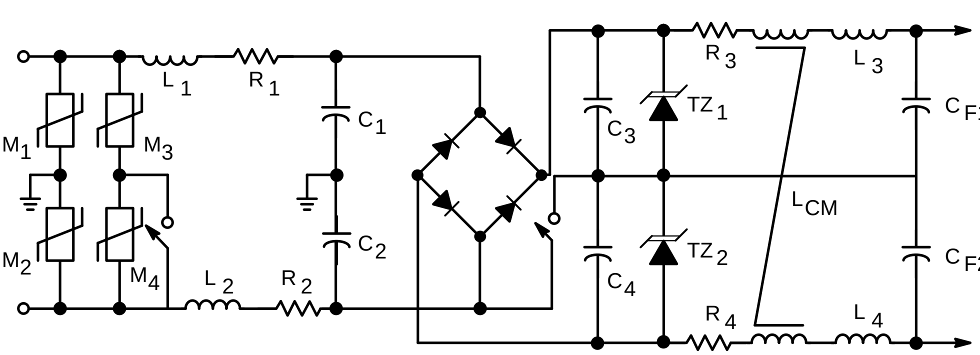


Figure 143. A Transient Suppression and EMI Filter Integrated with the Rectifier Circuit

Surge Current Protection

High current surges cause rapid heating of the semiconductor die. If reverse voltage is applied while the rectifier junction is above its rated temperature limit, a short may result from the combined stresses. Furthermore, even when little or no reverse voltage is present, excessive temperature can weaken the bond of the die to its package, causing thermal resistance to increase with time in service. A long term thermally regenerative process occurs, leading to diode destruction. In some cases, excessive current may cause the lead material between the rectifier terminal and the die to fuse.

The type and extent of protection needed is determined by the circuit characteristics and performance required. Circuit characteristics are roughly divided into two main types, and the extremes are:

- Type 1: Circuits with current limiting impedance
- Type 2: Circuits without current limiting impedance

These types are subdivided according to circuit performance requirements as to whether or not service may be interrupted upon failure of a diode. Since all circuits contain a certain amount of impedance, how a given circuit is classified depends on both circuit impedance and rectifier diode ratings. In general, a circuit is classified as Type 1 if the circuit impedance will keep fault currents below the single–cycle surge current rating of the rectifier diodes (I_{FSM}), and as Type 2 if the circuit impedance allows fault currents to go over the diode surge rating. Type 1 circuits may be protected with conventional fuses or circuit breakers chosen to open at currents below the surge current ratings of the rectifier diodes. Type 2 circuits, low impedances, are desirable from the standpoint of efficiency but require special fast–acting protective elements.

Current surges in rectifier circuits result from the following factors:

1. Capacitor in–rush
2. Direct current overload
3. Direct current short (fault)
4. Failure of a single rectifier diode

Capacitor in–rush is easy to observe and calculate and is discussed in Chapter 7; the other problems require some heroism in order to observe circuit conditions. Simulation of the fault may often be accomplished by using an SCR as a crowbar; the scope may be triggered by the same signal which fires the SCR.

Fusing Considerations

Basic operation of a fuse is as follows. When a current larger than normal is applied to a fuse, it starts to heat up significantly because power dissipation is proportional to the square of current. If this current continues, the fusible element or link eventually reaches its melting point. As the link melts, voltage builds up across the fusible material which causes arcing. Arcing continues until enough of the link has melted so that the applied voltage can no longer jump the gap. At this time current flow through the fuse stops. This basic operation is shown in Figure 144. The solid line shows the actual fault current due to the limiting action of the fuse and circuit impedance. The dashed line shows the available fault current that would flow without the fuse. As shown, melting of the fuse starts at point A. Typically, the current rises to a peak let–through current as shown at point B and then decays through the impedance of the arcing fuse to zero at point C. The time from point B to point C is the arcing time of the fuse. The melting time and the arcing time together make up the total clearing time of the fuse. Clearing time is often called blowing time.

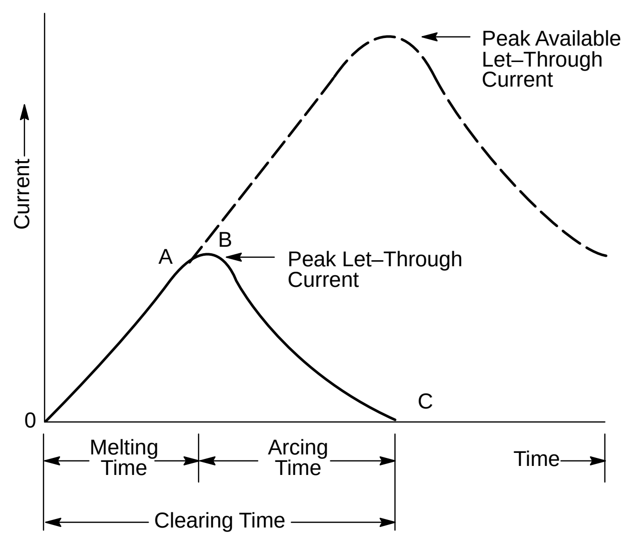


Figure 144. Action of a Fuse in a Circuit Subjected to a Current Surge

Three basic types of fuses are available: normal–acting or medium–acting, fast–acting, and slow–acting (often referred to as "slo–blo"). The different behavior of the fast–acting and the normal–acting fuse is caused by the material used for the link (this is the part of the fuse that melts) that results in the fast–acting fuse having a smaller thermal inertia. Similarly, the "slo–blo" fuse has a larger thermal inertia than the normal–acting fuse which results from use of a low–temperature melting point alloy and a heat sink. Figure 145 shows the general blowing time relationship for the three types at various overload conditions. Note the large blowing time required for the "slo–blo" compared to the other two. It is because of this difference that only medium–acting and preferably fast–acting fuses are used to protect rectifiers. The fast–acting fuses are sometimes called current–limiting fuses because of their fast blow times, although all fuses could be regarded as current limiters.

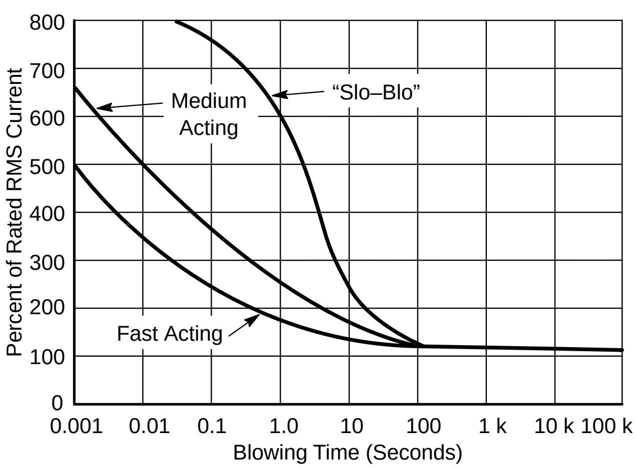


Figure 145. Chart Showing the Relation of Percent of Rated Current to Blowing Time for Typical Fuses

The best type of fuse in the fast–acting category is the silver–sand or semiconductor fuse. As the link (usually made of silver) melts, the arcing caused by the voltage developed across the fusible material is quenched by sand crystals placed around the silver link. The sand effectively lengthens the arc path. The particular size of the sand crystals is selected on the basis of the voltage and current ratings of the fuse. Fuses are available which limit fault current to as low as four times normal rated current.

Fuses are sensitive to changes in current and temperature not voltage. Figure 146 shows the effect changes in ambient temperature have on the current rating and blowing time. A fuse may be used at any voltage up to its maximum rating without any influence on the fuse's performance. Only after the link melting temperature is reached and arcing occurs in the circuit, are voltage and available power an influence on the fuse performance. It has been shown [5] that the clearing time I^2t (the effective energy that the fuse will let through) decreases as a fuse is used below its rated voltage.

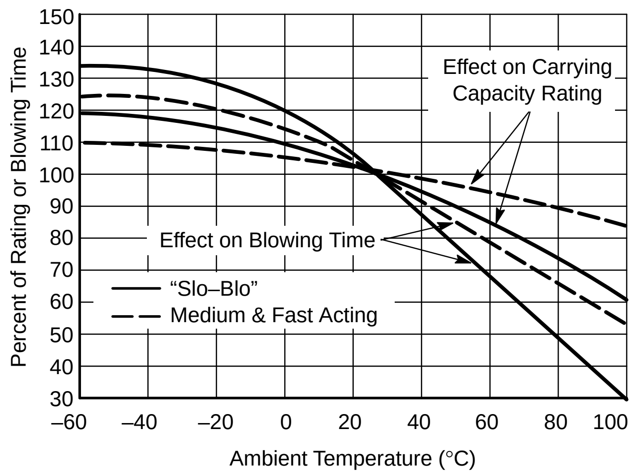


Figure 146. Action of a Fuse in a Circuit Subjected to a Current Surge

There are two common ways to select a fuse. One way is to compare the fuse and rectifier on the basis of surge currentthat is compare peak let–through current and rectifier surge current for a given fault current. Another way is to base the analysis on thermal energy, I^2t. It has been shown [6] that essentially equivalent heating occurs in a fuse and in a rectifier for a unit of I^2t where the I^2t of the fuse is based on a triangular surge current waveform and the I^2t of the rectifier is based on a sinusoid waveform. Based on this second approach, the following steps can be used as a guide for fuse selection:

1. Select a fuse with a current rating which is slightly higher than the current flowing through the rectifier under normal conditions. (Remember, this is an ac rms current rating.)
2. Check the voltage rating on the fuse to insure that it is equal to or greater than the maximum circuit voltage.
3. Estimate or measure the available fault current. Since it is circuit dependent, it can be found analytically by dividing the source voltage by the system impedance with the fault placed in the system.
4. With this information, the fuse tables are used to find the worst–case peak let–through current, I_{PLT}, and the total clearing energy, I^2t, for the selected fuse. A word of caution: fuse ratings are somewhat dependent upon mounting hardware. If a nonstandard mounting is used, consult the fuse manufacturer to see if there is any change in the fuse ratings.
5. Using the latter two parameters and assuming a triangular model for fuse current, find the total clearing time by using the formula:

$$t_C = 3I^2t/I^2_{PLT} \tag{9.1}$$

6. Obtain rectifier data required for this comparison (a plot of I^2t versus pulse base width). If surge current ratings, I_{FSM}, have been given instead, I^2t: can be calculated using the formula:

$$I^2t = I^2_{FSM}(t/2) \qquad (9.2)$$

The rectifier sub cycle I^2t ratings versus pulse width plot linearly on log–log paper.

7. Compare: With time as the independent variable, and knowing the I^2t for both the fuse and rectifier, it is now possible to determine if the rectifier is actually protected.
8. Put the fuse in the circuit, short the load or simulate the fault, and see if the rectifier is protected. Surge current should be examined on an oscilloscope.

EXAMPLE

Fault current = 3000 A
Rectifier I^2t = 27,0000 A^2s from 1 to 8.3 ms
Load fuse rating = 150 A (TYPE KAX)

For the 3000 A fault current, the total clearing I^2t and peak let–through current are found from Figures 147 and 148 to be 9500 A^2s and 2000 A. Using Equation 9.1, the clearing time is found as follows:

$$t_C = 3I^2t/I^2_{PLT} = (3)(9500)/(2000^2) = 7.1 \text{ ms}$$

Thus, the fuse will blow before the rectifier at a peak fault current of 2000 A. If the I^2t of the rectifier had not been given, it could have been found using the I_{FSM} rating.

I_{FSM} at 1 cycle is specified at 3600 A. Using Equation 9.2,

$$I^2t = I^2_{FSM}(t/2) = (3600)^2(8.3 \text{ ms})(10^{-3})/2 = 54 \times 10^3 A^2s$$

This value is twice the value of 27,000 given on the data sheet, thus showing that I^2t decreases by a factor of two going from a pulse width of 8.3 ms to 1 ms, a not unusual situation.

When circuit requirements are such that service is not to be interrupted in case of failure of a rectifier diode, the general approach is to use a number of diodes in parallel, each with its own fuse. Additional fuses or circuit breakers are used in the ac line and in series with the load for protection against dc load faults. Fuse values should be coordinated so that a diode fuse will open in the event of diode failure before the line fuses, the load fuses, or circuit breakers open. In the event of load faults, line or load elements should open before any of the diode fuses open. Such a system is illustrated in Figure 149.

Fuses are typically rated for use in ac circuits. When used in dc circuits, the voltage ratings are typically half of the ac rating. Some manufacturers offer fuses specifically designed for use in dc circuits, which are much more difficult to interrupt than ac circuits. There are subtle factors affecting fuse operation in dc circuits [9].

Redundant systems often need additional transient protection because fuse blowing causes a transient in the circuit upstream from the fuse. In the circuit of Figure 149, an MOV across all six rectifier banks is normally required.

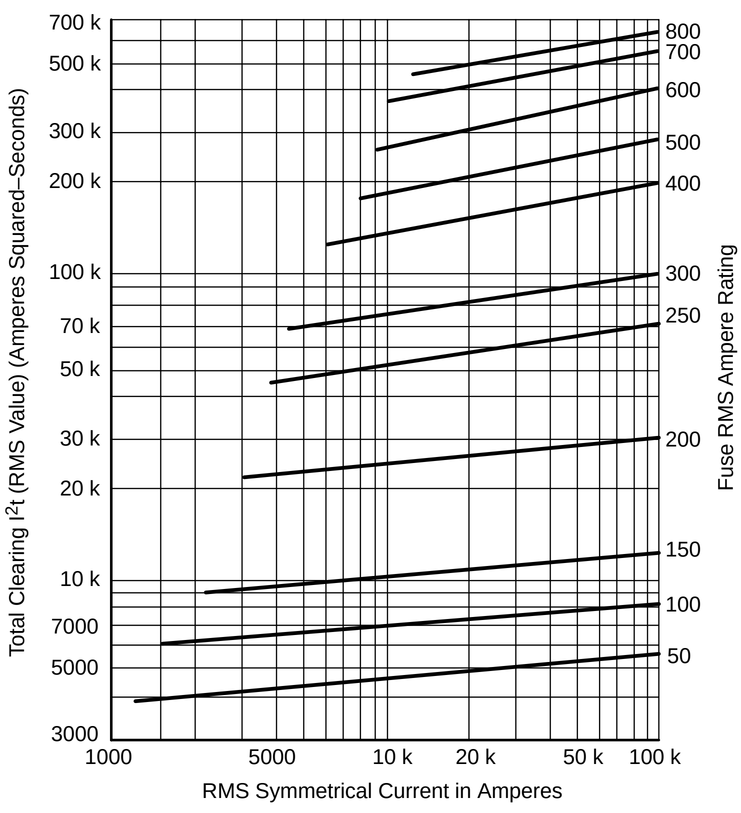


Figure 147. Total Cleaning Ampere Squared Seconds versus Available Fault Current, Type KAX Fuse

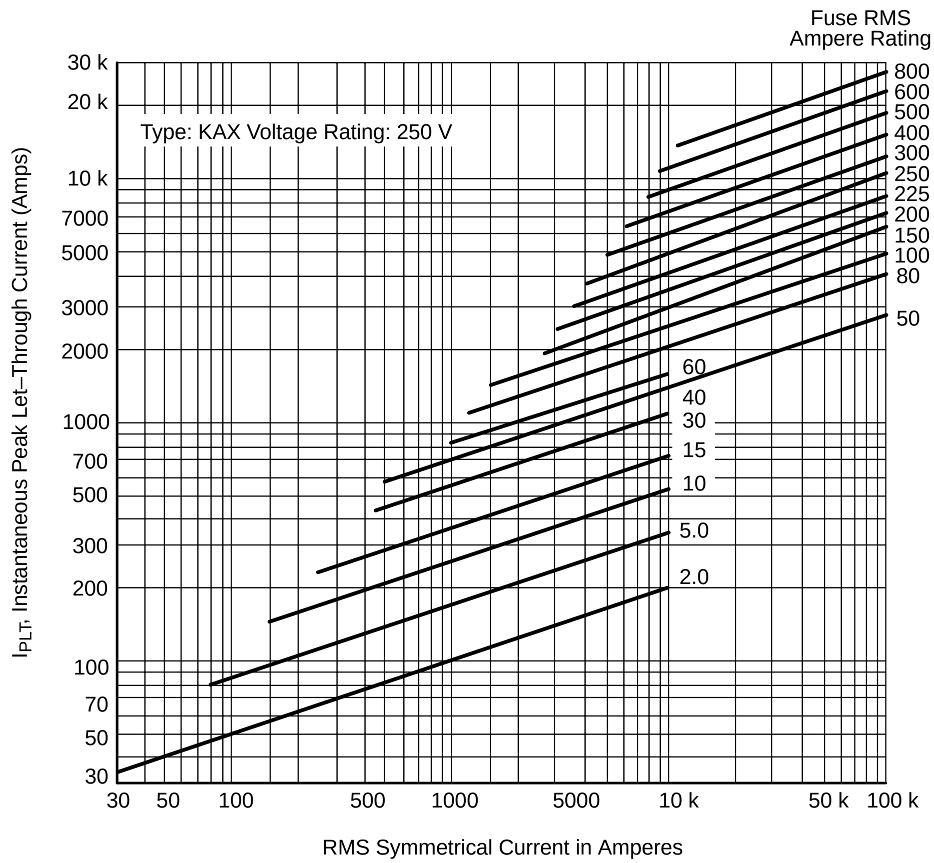


Figure 148. Peak Let–Through Current versus Available Fault Current, Type KAX Fuse

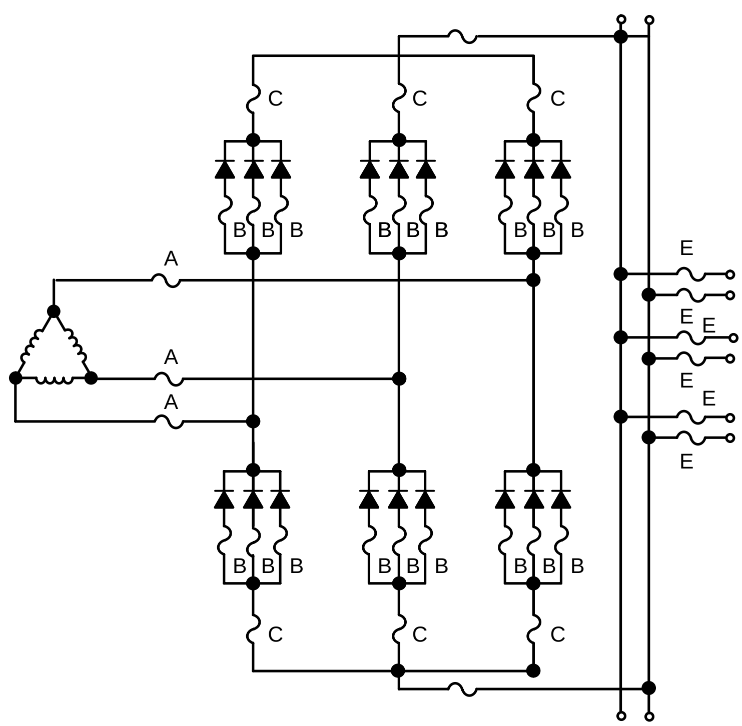


Figure 149. Fuse Placements in a Redundant System

References

1. IEEE Surge Protective Devices Committee, *Recommended Practice on Surge Voltages In Low–Voltage ac Power Circuits,* Institute of Electrical and Electronic Engineers.
2. Francois D. Martzloff and Thomas F. Leedy, *Selecting Varistor Clamping Voltage: Lower Is Not Better,* Proceedings of the 1989 Zurich International EMC Symposium, pp. 137–142.
3. Bill Roehr, *Hardening Power Supplies to Line Voltage Transients*, PCIM, June, 1986.
4. Bill Roehr, *An Effective Transient and Noise Barrier for Switching Power Supplies,* PCIM, Sept. 1986.
5. Martin Smith, *Some Considerations of Coordinating Fuses and Semiconductor*s, Bussman MFG. Co., Publication E–60–A–15.
6. Emil T. Schnoholzer, *Fuse Protection for Power Thyristors,* IEEE Transactions on Industry Applications. Vol. lA–B, #3, May/June 1972.
7. Bob Botos and Bob Haver, *A Fuse–Thyristor Coordination Primer*, ON Semiconductor Application Note AN568, ON Semiconductor, Phoenix, Arizona.
8. F. B. Golden, *Take the Guesswork out of Fuse Selection*, The Electronic Engineer, July 1969. pp. 71–75.
9. A. F. Howe, P. G. Newberry, N.P.M Nurse, *DC Fusing in Semiconductor Circuits,* IEEE Transactions on Industry Applications, Vol 1A–22, No. 3, 1986.

Chapter 10

Basic Diode Functions in Power Electronics

Basic Diode Functions in Power Electronics

The buck and boost converters are generally regarded as the basic building blocks for power conversion [1]. By combining these basic building blocks with some kind of transformation function, many new complex and all known converters are derived.

The basic topologies of buck and boost converters are shown in Figures 150 and 151. In operation, the transistor switch, often referred to as the power switch, is alternately turned on or off. When the power switch is on, the amount of energy stored in the inductor is determined. When the power switch is off, some or all of the stored energy flows to the load and the filter capacitor, C_F. These four essential elements–power switch, diode, inductor or choke, and capacitor–play key roles in all power conversion circuits. Despite its essential role, the diode has received far less attention in the literature than its more glamorous associate, the power switch.

Buck and boost converters are designed to operate in one of two modes. In “continuous mode” operation, the inductor or choke current never becomes zero during the switching cycle. In “discontinuous mode” operation, the choke current decays to zero, often called “drying out,” before the power switch turns on again. The overall characteristics of the converter and the requirements placed upon the four essential elements differ greatly depending upon whether operation is confined to the continuous or discontinuous mode.

- Naturally commutated or soft switching circuits
- Forced commutated or hard switching circuits

In naturally commutated circuits, diode turn off is primarily by internal recombination and no appreciable reverse current flows. Basic circuit functions that are normally naturally commutated include:

- Rectification, discontinuous mode
- Catching
- Clipping
- Clamping

To maintain soft switching operation after forward current ceases, sufficient time for stored charge to recombine must elapse before the circuit requires the diode to be a high impedance. Diodes must be chosen so that their “time constant” (lifetime, $*T_S$) is less than 20% of the time allocated during the pulse repetition period. Duty cycles vary over a wide range in pulse–width–modulated (PWM) circuits so that the time for diode turn off may be below 10% of the pulse repetition period in some designs.

Forced commutation occurs in circuits where appreciable forward current is flowing when a change in circuit state forces the diode to turn off. Usually a high reverse current occurs during turn–off which must be handled by other circuit elements.

Basic circuit functions that commonly use forced commutation include:

- Freewheeling
- Rectification, continuous mode

In hard switching circuits, the closer the diode is to the ideal one with no stored charge, the better the circuit performs. As will be shown later in a detailed analysis, the low–impedance state which the diode maintains during the t_a portion of reverse recovery time causes high stress on other circuit elements which can be mitigated by using a fast recovery diode.

Circuits used in power electronics generally can be placed in one of two categories as far as the demands imposed upon the diode are concerned. They are:

In many circuits, the element that suffers the highest stress is a power transistor switch which normally is the most highly stressed and failure–prone component in a power electronics system. Consequently, any means to reduce stress on the power switch is helpful to overall system reliability and efficiency. Choosing a fast diode is one means, and using a transistor snubber is another. Snubber circuits, discussed in the next chapter, usually employ diodes which also need to be selected for proper operation.

This chapter examines the basic diode functions separately with the objective of identifying the requirements of the diodes. Used in conjunction with additional active and passive components, the basic functions are embedded in a number of widely used sub–circuits that play important roles in power electronic systems. Some practical concerns such as power switch drive circuit problems, parasitic inductance caused by layout, and use of a diode snubber are discussed near the end of this chapter. The next chapter shows some commonly used power electronic circuits to illustrate the principles of diode selection developed in this chapter.

Rectification, Discontinuous Mode

The discontinuous mode boost converter shown in Figure 150 is a common example of rectification where the switching is soft. Waveforms are shown in Figure 152 for an idealized case where resistance in the circuit is zero. However, since resistance causes power loss and heating, considerable effort is spent in power conversion circuits to minimize resistance and waveforms in practice are not significantly different from those shown.

When the power switch Q_S is turned on, current ramps up in Q_S and the choke stores energy in its core. During this time, shown from t_0 to t_1, the circuit voltages are fixed. The transistor on–state level is nearly zero, the inductor has V_I impressed across it and the diode must block the output voltage. When the transistor turns off, the voltage across the transistor reverses and the "flys back" according to the relationship, v = –L di/dt. When the voltage across the power switch starts to exceed the output voltage V_0,the diode conducts allowing the choke energy to flow into the load and the capacitor.

The output voltage is a complex function of the switch duty cycle, the load current, and the output filter capacitor but it must exceed the input voltage. An attempt to have $V_0 < V_1$ would cause dc current to flow from V_1 and the circuit would not function. Because of the flyback action of the choke voltage, the converter's sometimes called a flyback converter, but this term is more properly reserved for a popular transformer–coupled converter that can be shown to be a cascade connection of a buck and a boost converter. The boost circuit is also called a ringing choke or step–up converter.

The diode requirements are not stringent in this application. For low loss, the forward voltage drop and dynamic impedance during turn–on should be low. The diode only needs to block V_0. The carrier lifetime $*T_S$ should be no more than 20% of the charging time of the capacitor for the boundary condition where the time from t_3 to t_4 is zero. Otherwise its conduction would extend into the time when the switch turns on, resulting in reverse recovery losses.

When the input voltage is first applied, the transistor is off. An inrush current flows to charge the filter capacitor, limited only by the characteristics of the LC network and any series resistance in the wiring and components. The current wave is a one–half cycle of a damped sinusoid and its peak value can be several times the steady state value that the components normally experience. Either the components must be sized to handle the inrush or means must be taken to limit its magnitude.

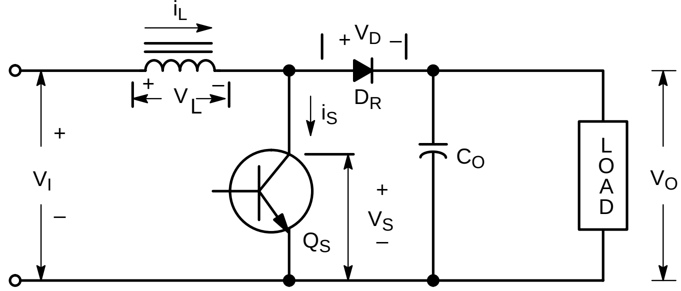


Figure 150. Output Section of a Boost Converter; the Diode Performs Rectification

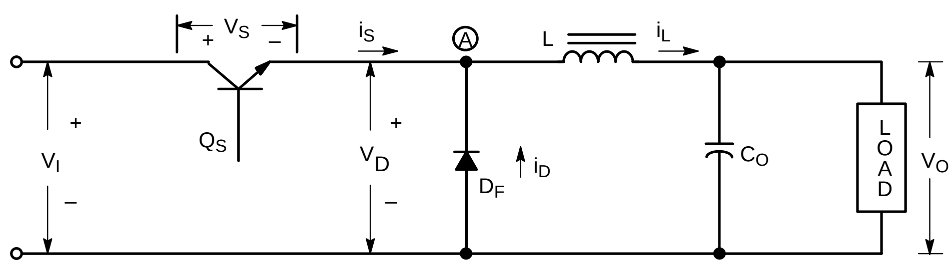


Figure 151. Output Section of a Buck Converter; the Diode Performs a Catch Function

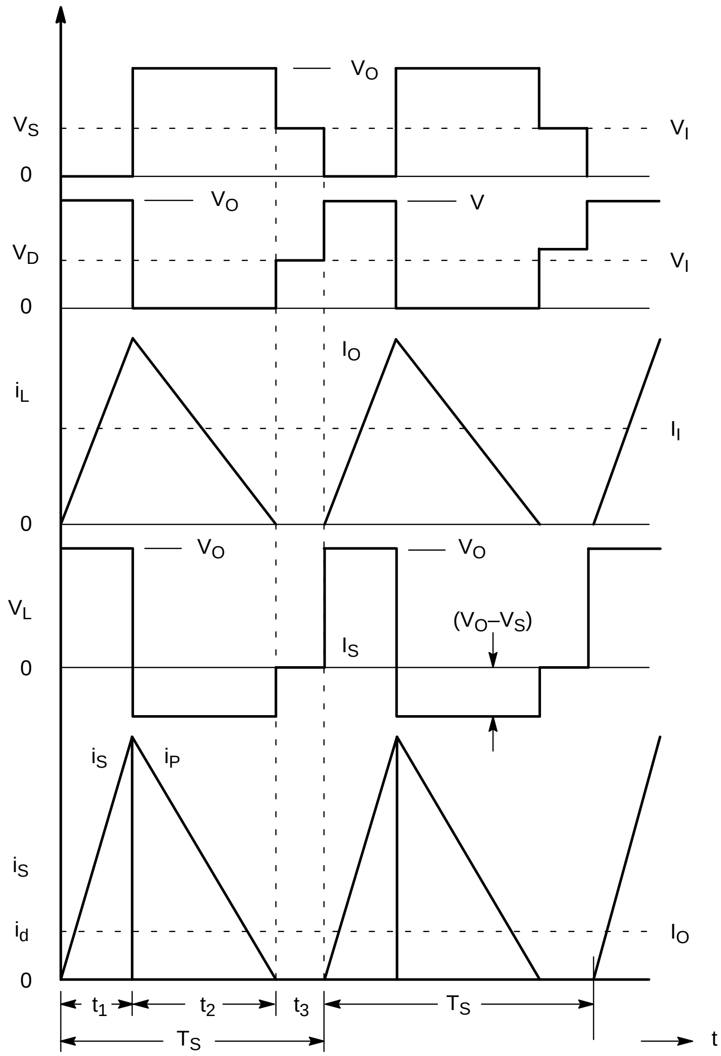


Figure 152. Idealized Waveforms for Boost Converter Operating in the Discontinuous Mode

Catching

The discontinuous mode buck converter is sometimes used at low power levels. The diode performs a catch function rather than a rectification function. Waveforms are shown in Figure 153, idealized to the extent that resistances in the components are neglected.

When the power switch is turned on, current ramps up in the inductor; charging the output capacitor and delivering current to the load. During transistor turn–off, which begins at time t_1, the inductor causes the voltage at point A to go negative. It is prevented from dropping below the reference level by the diode. The waveform is sometimes said to be "caught."

Without the catch diode, the voltage across the switch could reach destructive levels. The transistor current now flows through the diode which permits the choke to continue supplying current to the load. Since the diode maintains current flow during pan of the cycle, it is often called a "freewheeling" diode. By maintaining current flow when the transistor is off, circuit efficiency is improved.

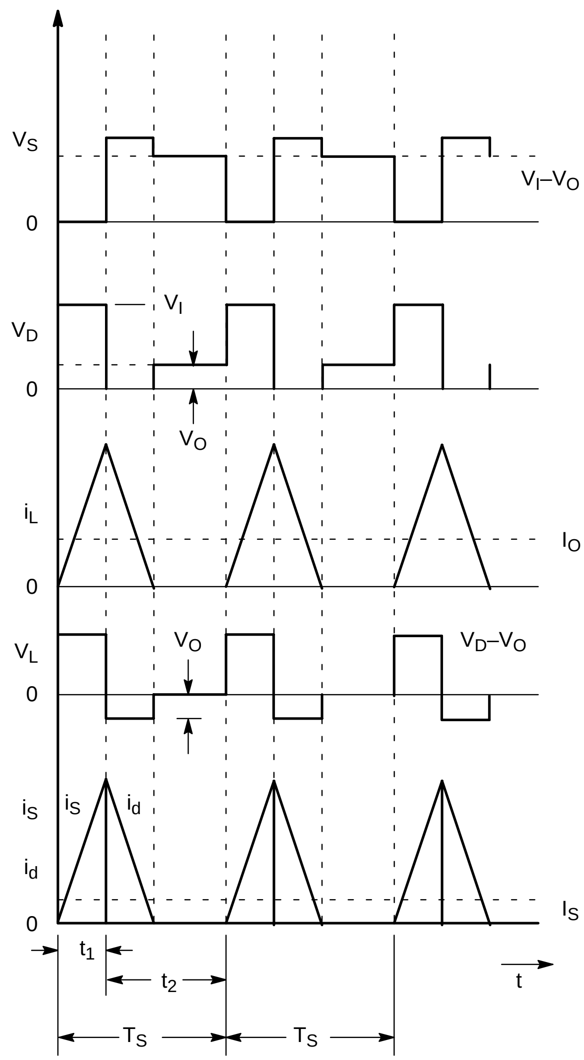


Figure 153. Idealized Waveforms for Buck Converter Operating in the Discontinuous Mode

Mitigating inductive kick is a common problem in power electronics. The push–pull circuit of Figure 154 supplies power to a resistive load via a transformer. The inductance L_P represents the sum of parasitic wiring inductance plus transformer leakage inductance.

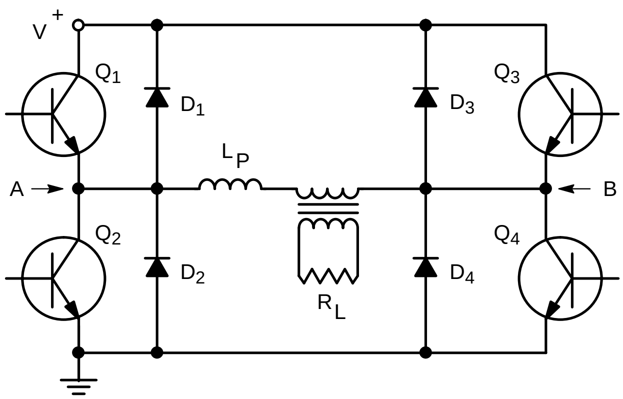


Figure 154. A Push–Pull Full Bridge Circuit Using Catch Diodes

Assuming very fast switching, when Q_2 and Q_3 turn on at time t_1, V^+ appears at point B and point A is pulled to ground. Current builds up at a rate determined by the circuit time constant. When the signal to turn off the transistors is received, they will not turn off at the same time because of their storage delay time. Suppose Q_2 turns off first. It will need to block V^+ plus the inductive kick generated by L_P, if the catch diodes are not in the circuit. A similar situation occurs when transistors Q_1 and Q_2 are switched.

The inductive kick is given by Faraday's law as $V_L = –L_P\, di/dt$. It is instructive to calculate the inductive kick for the typical values in a high–frequency power electronic circuit. Assume that $V^+ = 400$ V. $R_L = 20\ \Omega$ (current = 20 A) and $L_P = 0.2\ \mu$H. If the switch opens in 20 ns,

$$V_L = -(0.2)(10^{-6})(20)/(0.02)(110^{-6}) = -200\ \text{V}$$

Raising the transistor switch blocking voltage requirement imposes penalties in terms of cost and performance. Consequently, it is common practice to use catch diodes to confine the voltage across the switches to values between the rail voltage and ground. The diodes also prevent the current in the inductor from changing quickly. The current is commutated from an off–going switch to a diode and decays according to the relationship, $\Delta t = IL/V$.

The diode forward recovery characteristic is of utmost importance in this application since it determines how far above the rail or below ground the inductive kick can drive the voltage. Since the diode is on only briefly, other characteristics are relatively unimportant except, of course, the diode must block the voltage V^+. As with other soft switching circuits, when used at high frequencies the diode lifetime should be less than a fifth of the time between pulses. Otherwise, a reverse current will flow through the diode, which must be handled by the power switch at turn–on.

Clipping

In clipping, a portion of a wave is flattened off or limited to some arbitrary level, irrespective of the amplitude of the original signal. Peak clippers find use in power electronics to protect semiconductors from overvoltage.

A peak clipper, also called a peak limiter, prevents the positive (or negative) amplitude of the wave from significantly exceeding a value set by the clipper. A simple example of such a clipper is illustrated in Figure 155. When the instantaneous value of the input voltage wave lies between V_A and V_B, neither of the diodes conduct, and the input wave is transmitted directly to the output terminals without change. On the other hand, when the input voltage is more positive than V_A, diode D_A will conduct and thus prevent the output voltage from rising appreciably above V_A. Similarly, when the input Voltage becomes more negative than V_B, diode D_B will conduct and clip the negative peaks of the output voltage at a level approximately equal to V_B. In circuits of this nature, the output is often said to be "clamped" to V_A or V_B. However, a clamping circuit does not change the waveshape or limit its amplitude, as described in the next section. For the diode clippers to be effective, the series resistance R_S must be considerably greater than the forward resistance R_F of the diodes, since R_S and R_F form a voltage divider. In addition, the turn–on impedance should be low to minimize overshoot and the diode lifetime fast enough to follow the input waveform.

If the batteries in Figure 155(a) are replaced by large capacitors, the circuit functions as a dynamic clipper when driven by a repetitive ac wave. The left capacitor is charged to the positive peak of the ac wave and the right capacitor is charged to the negative peak. Should a transient exist on the wave in excess of the normal peak value of the wave, it will be clipped. The action is essentially the same as when voltage sources are used, except the sources adjust to the peak of the ac wave. To permit the voltage level to drop in response to a change in the amplitude of the wave, a bleeder resistor across each capacitor is needed. The RC time constant is usually chosen to be about five times the period of the input wave.

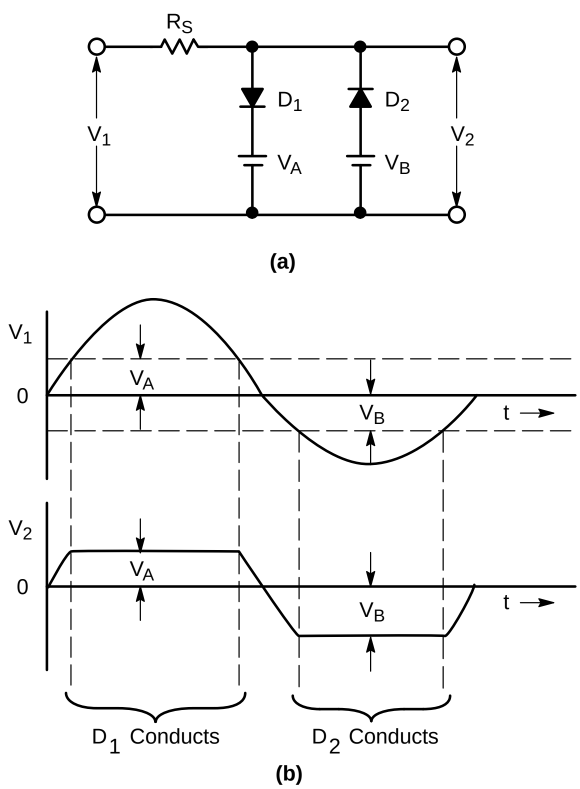


Figure 155. Diode Clipping Circuit and Waveforms when Peak Clipping a Sine Wave: (a) Circuit; (b) Waveforms

Clamping

Many asymmetrical waveforms possess a dc reference component which remains fixed–it is not a function of the waveform. When such a waveform is passed through a transformer or a capacitor, the dc component is not transmitted. Loss of the dc reference is often a problem because the absolute value of the positive and negative portion of the wave is dependent upon the duty cycle which usually varies over a wide range in a power converter. In order to properly drive a power stage when the drive waveform is coupled through a transformer, the positive and negative peaks must remain at fixed levels. To regain the original waveform it becomes necessary to reinsert a dc component of appropriate amplitude. This process is termed dc restoration. It is a special example of a more general process known as clamping, in which a reference level is introduced that has some desired relationship to the wave.

The simplest form of a clamping circuit consists of a diode associated with a resistance–capacitance network, as illustrated in Figure 156. The resistance R in these circuits is relatively large, and the time constant RC is considerably greater than the time of one cycle of the applied wave. If circuits are used that insert a dc component of such amplitude and polarity that the negative peaks of the original wave are at zero level, as shown in Figure 156(a), then it is said that the negative peaks of this wave are clamped to zero. The same wave with its positive peaks clamped to zero volts and to $+V$ volts is shown in parts (b) and (c), respectively.

The detailed action of the diode clamp of Figure 156(b) can be understood by considering Figure 157. If RC is so large that it can be considered infinite, then the diode charges capacitance C to a voltage V' that just barely prevents the output voltage V_0 from going positive; thus, V' is the reinserted dc voltage. However, if RC is high but not infinite, then some of the charge on the capacitor leaks off during the time interval $t_3 - t_2$. This causes the output wave to decay with a time constant RC during the time interval $t_3 - t_2$; however, if RC is very much greater than the time $t_3 - t_2$, the change ΔV_0 in the amplitude of the output voltage will be small. At t_3 the voltage at the output terminals increases suddenly in amplitude by the peak–to–peak amplitude V of the applied wave and makes the output voltage go ΔV_0 volts positive at time t_3. This causes current to flow into the capacitor to replace the charge that leaked off during the previous period, as shown. The time constant associated with this current flow corresponds to C being charged through the diode and source resistance.

In common with the diode requirements of other soft commutated circuits, diode forward recovery should be fast and the lifetime fast enough to not have any stored charge left in the diode when the next cycle of the input wave occurs. The proper value for the time constant RC is a compromise between two conflicting requirements. The larger this time constant, the smaller will be the variation ΔV_0 of the clamping voltage during the cycle, and, hence, the smaller will be the positive spike at t_1 and t_3. On the other hand, the larger the RC product, the more slowly is the clamping voltage able to adjust itself to reductions in duty cycle or amplitude of the wave that is to be clamped.

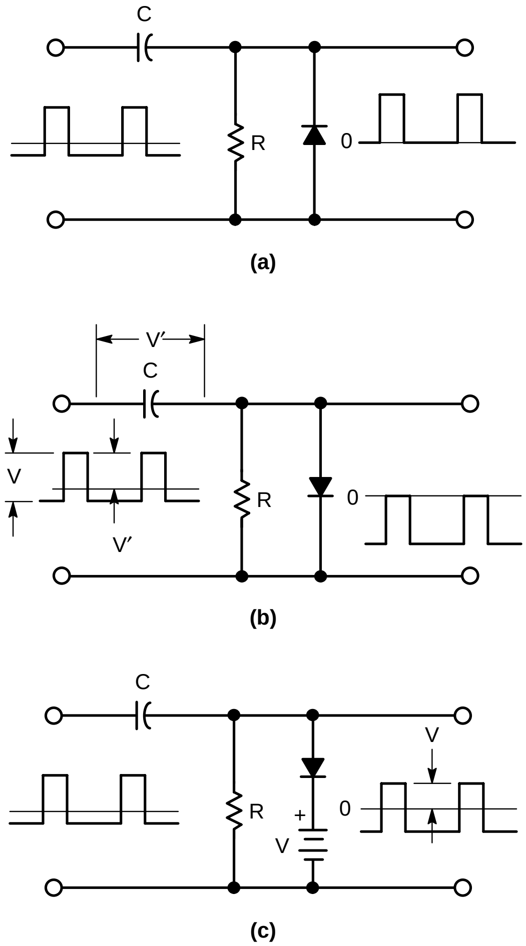


Figure 156. Clamping Circuits and Waveforms; (a) Negative Peaks Clamped to Zero; (b) Positive Peaks Clamped to Zero; (c) Positive Peaks Clamped to V

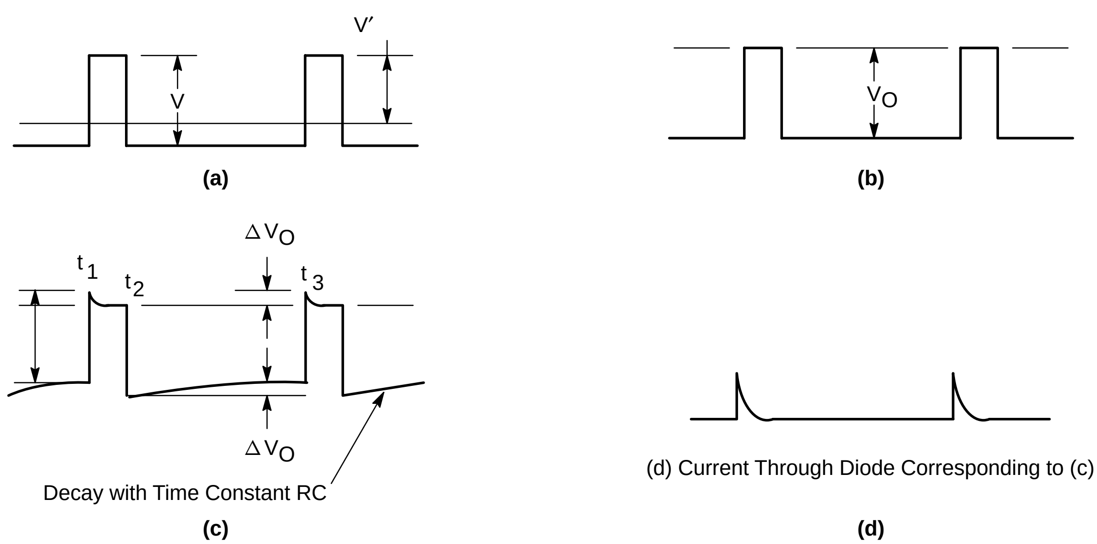


Figure 157. Detailed Waveforms for the Clamping Circuit of Figure 156(b): (a) Applied Wave; (b) Output Wave–RC Infinite; (c) Output Wave–Moderate RC; (d) Current through Diode Corresponding to (c)

Freewheeling

An important use of diodes is to keep current flowing in an inductor when the source of energy is pulsed off. When used in this manner, the diode is called a freewheeling diode. Action is similar to the catch diode used in the discontinuous mode buck converter, but the diode is carrying forward current when the power switch turns on. A common application is the buck converter of Figure 151 operating in the continuous mode. In this case, inductor L_F is chosen large enough so that its current never goes to zero.

Waveforms for the buck converter operating in the continuous mode are shown in Figure 158 They are drawn to the same scale as those shown in Figure 153, which apply to the discontinuous mode. It is quite apparent that continuous mode operation places much less peak current demand upon the switch, diode, and inductor. Although not shown, the ripple current through the filter capacitor is also much less. Consequently, the current capabilities of the key power components can be downsized from those required by the discontinuous mode.

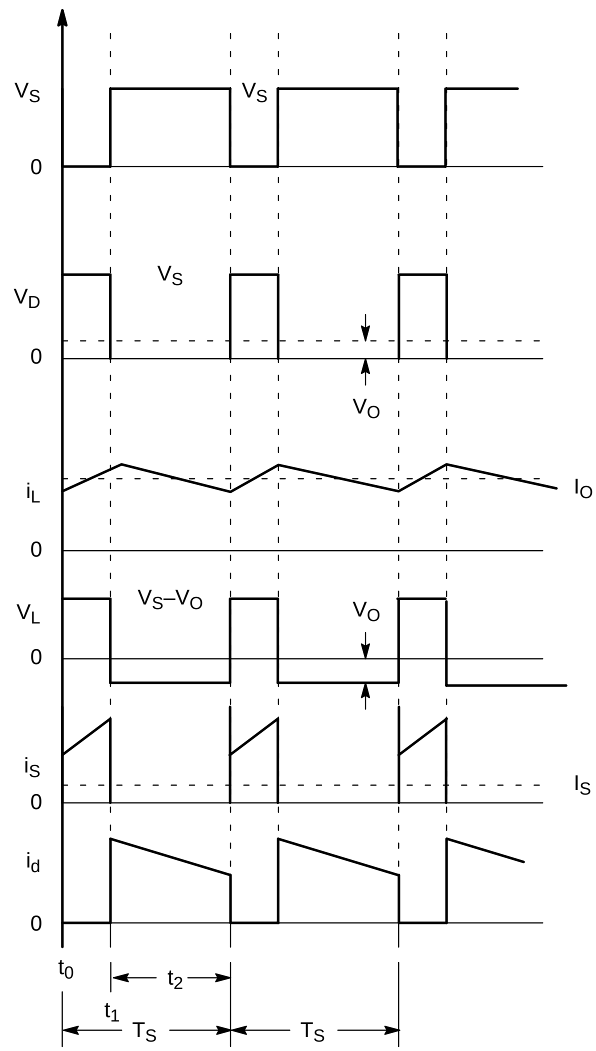


Figure 158. Idealized Waveforms for a Buck Converter Operating in the Continuous Mode

There are two disadvantages of continuous mode operation. The inductance of the choke must be much larger and the diode must experience forced commutation when the power switch turns on. The resulting reverse recovery current adds a current spike to the turn–on waveform which reduces circuit efficiency.

Waveforms for the boost converter operating in the continuous conduction mode are shown in Figure 159. Continuous mode operation offers the same reduction of peak currents in the components of the boost converter as occurs in the buck converter and it introduces the same problem, forced commutation. Accordingly, low stored charge is a desirable diode attribute.

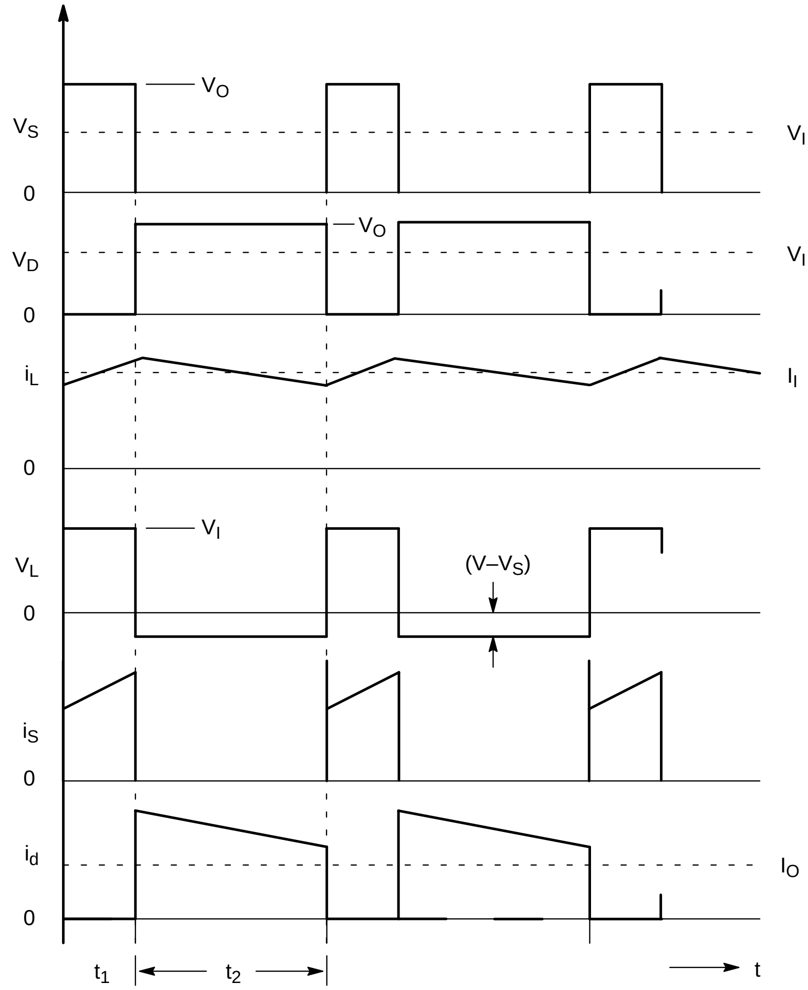


Figure 159. Idealized Waveforms for a Buck Converter Operating in the Continuous Mode

The rectifier diode (D_R) has a dual role. It is a rectifier since it passes intermittent pulses of current to the filter capacitor, C_0. It also is a freewheeling diode since it maintains the flow of choke current when the transistor switch is off.

To provide high efficiency, the diode D_F must meet stringent requirements. Since it supports the full load current possibly up to 90% of the time, it should have a low forward drop. When the switch turns off, a low dynamic V_F is required for low loss and to minimize the switch voltage stress. (At this instant the freewheeling diode also acts as a catch diode; therefore, the terms “freewheeling” and “catch” are often used interchangeably.) When the switch turns on, both the switch and the diode are subjected to a high current, limited only by the circuit parasitic inductance and diode reverse recovery time. Consequently, low diode stored charge is a very important asset, not only for high diode efficiency, but also for high switch efficiency and low stress.

Rectification, Continuous Mode

In the basic buck and boost converters, both the power switch and the diode participate in producing a dc output from a dc input. A rectifier function as classically viewed– that is, changing an ac wave to a unidirectional wave–does not exist. However, when a transformer is present, the diode that functions as a rectifier becomes obvious.

Countless numbers of converters [2] can be derived by combining buck and/or boost topologies with transformers and chokes. One of the more popular topologies is the forward converter shown in Figure 160. It normally operates in the continuous mode. A more detailed analysis of the forward converter is given in the next chapter; at this point the rectification function is the focus of the discussion. When the power switch is on, the dotted end of the secondary wiring is positive; the rectifier diode D_R conducts, permitting current to be transmitted from the input source to

the choke and the load. When the power switch is off, the transformer polarities reverse and output current is commutated to the freewheeling diode while D_R blocks. When the switch turns on again, current is commutated from the freewheeling diode to the rectifier diode. Both changes of state occur under forced commutation conditions and therefore efficiency is improved as the diode stored charge approaches the ideal of zero.

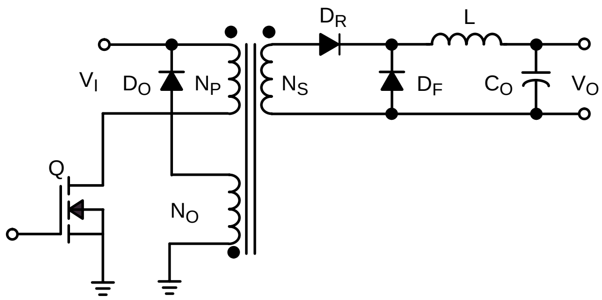


Figure 160. Output Section of a Forward Converter

Forced Commutation Analysis

The previous material has pointed out several power supply circuits where a power transistor must clear diode stored charge while turning on. The equivalent circuit of Figure 161. is useful for analyzing the commutating interval for any of these hard switching circuits. The essential common features are that during the switching interval the transistor must conduct the current formerly flowing in the diode, and the voltage across the transistor must drop from a relatively high off–state level to an on–state level that approaches zero. To show the major impact of diode stored charge, an in–depth analysis of the switching process follows. The analysis reveals numerous advantages of using diodes having a low stored charge–which is synonymous with fast reverse recovery–in forced commutation circuits.

The current remains constant as current is switched between the transistor and diode. At any instant of time, the sum of transistor current and diode current equals the inductor current, modeled as a current source. The pertinent waveforms are shown in Figure 162. The solid lines apply for a practical diode, while the dotted lines show behavior for an ideal diode having no switching losses.

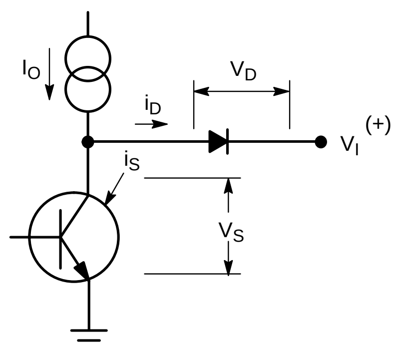


Figure 161. Equivalent Circuit Used for Analyzing Losses in Forced Commutated Circuits

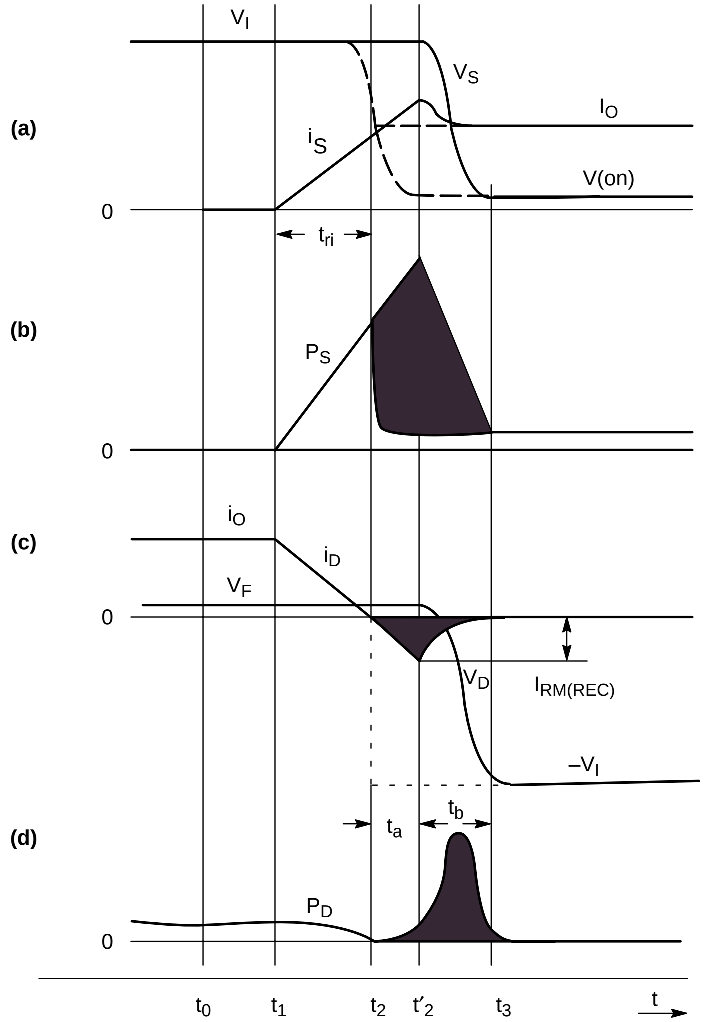


Figure 162. Forced Commutation Waveforms: (a) Power Switch Voltage and Current; (b) Power Switch Dissipation; (c) Diode Voltage and Current; (d) Diode Dissipation. Diode Forward Drop and Switch On–State Levels Are Exaggerated

At the start of the cycle at time t_0, the diode is conducting and the transistor is off. As the transistor turns on at t_1, current is diverted from the diode. However, the diode is still a low impedance and therefore the transistor must withstand the full input voltage as its current builds up. At t_2, all the inductor current is assumed by the transistor. If the diode were ideal, it would turn off and the commutation would be complete. The dashed lines indicate what the waveforms would be if the diode were ideal, that is, if it did not have stored charge or junction capacitance. With a real diode, the low–impedance state is maintained during the t_a interval of diode reverse recovery. Thus, the transistor current overshoots because of the reverse recovery current. During the diode t_a time, shown on Figure 162. as the interval from t_2 to t'_2, the transistor is subjected to high stress. During the diode t_b time, shown on Figure 162. as time interval t'_2 to t_3, both the transistor and the diode are subjected to high stress. During the diode t_b interval, diode current decays to zero and the transistor current mirrors it, dropping to its on–state value. Also, the transistor voltage drops to its saturation level while the diode voltage rises to the supply voltage.

The power dissipation waveforms for both the transistor and diode are also shown in Figure 162. The shaded area on the transistor waveform indicates the dissipation caused by the diode stored charge and capacitance. Note that the dissipation during t_a is almost as much as the dissipation during the switching of inductor current to the transistor in the time interval t_1 to t_2. This situation is not uncommon when the diode reverse recovery time is approximately equal to the transistor rise time. Reverse recovery current raises transistor dissipation appreciably because it occurs when the transistor is sustaining its full off–state voltage.

By approximating the voltage and current change with straight lines, the transistor dissipation during turn–on can be found from the waveforms using geometry. It is given by:

$$P_T = \left(\frac{V_I I_O}{2}\right)\left(\frac{t_{ri}}{T_S}\right) + V_I\left(I_O + \frac{I_{RM(REC)}}{2}\right)\left(\frac{t_a}{T_S}\right) + \left(\frac{V_I I_O}{2}\right)\left(\frac{t_{rv}}{T_S}\right) + \left(\frac{V_I I_{RM(REC)}}{4}\right)\left(\frac{t_b}{T_S}\right) \quad (10.1)$$

The first term is the dissipation in the transistor during current rise time if the diode were ideal. The second term has two parts: the first part is caused by the lengthening of transistor turn–on by the diode t_a interval and the second part accounts for the overshoot current caused by t_a. The third term is the dissipation during transistor voltage fall time if the diode were ideal. The last term is dissipation produced during the t_b portion of diode recovery. It assumes that $t_b = t_{rv}$, which is not necessarily true; however, it is often true that the t_b contribution to transistor dissipation is comparably small.

Nearly all the diode dissipation occurs during time t_b. It is difficult to quantify since the voltage during the t_b interval is also affected by the transistor and circuit inductance. Additionally, the t_b interval is a function of the diode capacitance which produces reactive power not contributing to its power dissipation. For a slow diode, where t_b is much longer than t_{rv}, the diode dissipation may be approximated by assuming that the full reverse voltage is across the diode during t_b. In this case the transistor dissipation during t_b is zero, and the diode dissipation is:

$$P_D = V_I\left(\frac{I_{RM(REC)}}{2}\right)\left(\frac{t_b}{*\,T_S}\right) \quad (10.2)$$

Regardless of how power is shared during the t_b interval, Equation 10.2 is an accurate expression for the sum of the transistor and diode dissipation. When the diode is fast, t_b will be much less than t_{rv} and dissipation in the diode will be very small since the voltage across it will not have time to rise appreciably. Consequently, when fast diodes are used, the contribution of t_b, to dissipation in both the transistor and diode is small.

The first two terms in Equation 10.1 quantify the power dissipation during the current rise time interval. The effect of the diode recovery current during the t_a interval is found by expressing the first two terms with $I_{RM(REC)}$ as a variable instead of t_a. From the waveforms of Figure 162, observe that the current waves have a constant slope from t_1 to t'_2, therefore, $t_a/t_{ri} = I_{RM(REC)}/I_O$. Substituting for t_a in Equation 10.1 and after some manipulation, the following expression for the dissipation P_I, during the t_{ri} and t_a intervals is derived:

$$P_I = \frac{V_I I_O}{2} \cdot \frac{t_{ri}}{T_S}\left[1 + \frac{2I_{RM(REC)}}{I_O} + \frac{I^2_{RM(REC)}}{I^2_O}\right] \quad (10.3)$$

The expression outside the brackets in Equation 10.3 is the transistor dissipation when the diode is ideal–that is, $I_{RM(REC)}$ is zero. The term in brackets is a multiplying factor (P_R) which shows the ratio of P_I with an ideal diode to P_I with a real diode during the current commutation interval. Because of the squared term, the ratio increases very rapidly with increasing $I_{RM(REC)}$. Figure 163. illustrates the heavy penalty that $I_{RM(REC)}$ imposes upon the power transistor. When $I_{RM(REC)} = I_O$, P_I is four times that occurring with an ideal diode.

It is important to recognize that Equation 10.3 is valid for all types of diodes. It does not matter whether the source of $I_{RM(REC)}$ is stored charge in a junction diode or capacitance in a Schottky diode; high $I_{RM(REC)}$ causes a severe increase in the transistor dissipation during turn–on.

Equation 10.3 can be stated in terms of t_{ri} to show the effect of changing the transistor current rise time. From Figure 162. the recovered charge (Q_a) during time interval t_a is found by geometry to be:

$$Q_a = t_a I_{RM(REC)}/2$$

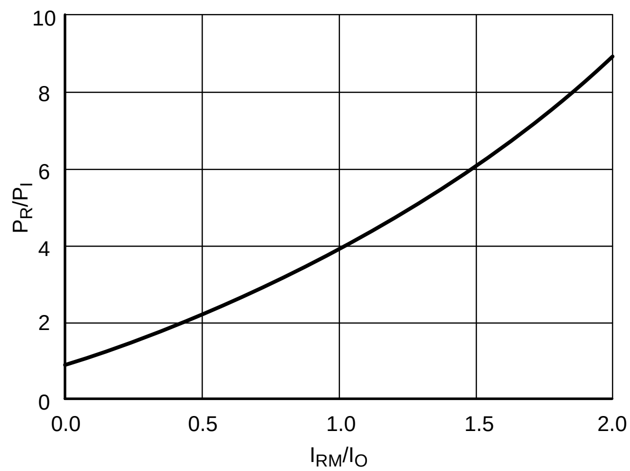


Figure 163. Increase in Switch Dissipation during Current Commutation Caused by Diode Recovery Current

Solving for t_a and substituting into Equation 10.3, the following expression is derived after some algebraic manipulation:

$$P_I = V_I I_O \frac{T_{ri}}{2T_S} + V_I \frac{Q_a}{T_S} + \left(\frac{2V_I}{T_S}\right)\sqrt{Q_a t_{ri} I_O} \quad (10.4)$$

The first term is a dissipation component that is proportional to transistor rise time. It is the dissipation that occurs when the diode is ideal: that is, $Q_a = 0$. The second component is fixed by the diode charge while the third term is affected by both diode charge and transistor rise time. It is evident that a point of diminishing returns is reached where improvements in rise time cause negligible reduction in power dissipation. The primary reason is that as t_{ri} is reduced, $I_{RM(REC)}$ increases. Secondarily, as t_{ri} (diode di/dt) is reduced, Q_a increases somewhat in a practical diode.

Equation 10.4 can be put into another form to access the point of diminishing returns by malting some simple approximations. For a fast pn junction diode, the charge (Q_a) which appears during time interval t_a is approximately equal to the stored charge Q_S when di/dt is rapid. Also the forward

diode current equals the transistor current which equals the load current. Therefore,

$$Q_a \sim Q_S \sim I_F \cdot T_S \sim I_C \cdot T_S \sim I_O \cdot T_S$$

Substituting into Equation 10.4 and performing some algebraic manipulations, the following expression is formed:

$$P_T = \left(\frac{V_I I_O}{2}\right)\left(\frac{t_{ri}}{T_S}\right)\left[1 + \left(\frac{2 * T_S}{t_{ri}}\right) + \sqrt{\frac{16 * T_S}{t_{ri}}}\right] \quad (10.5)$$

The term outside the bracket is by now the familiar expression for transistor dissipation when an ideal diode is used ($*T_S = 0$). The bracketed term shows the factor causing increased dissipation in terms of the ratio of the diode lifetime to the transistor current rise time. This term P_R is plotted in Figure 164 and offers insight for choosing an appropriate diode for use in a particular application.

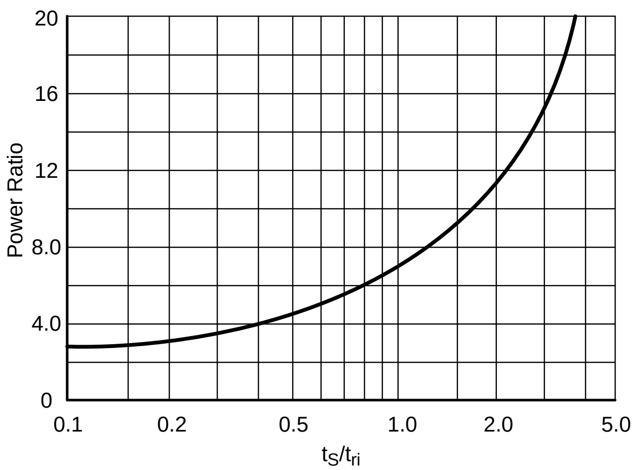


Figure 164. Effect of Diode Lifetime on Power Dissipation during Current Commutation

The curve asymptotes to a little over two when $*T_S$ is much less than t_{ri}, which is approaching the condition when the diode is ideal. However, the ratio is only four when $*T_S \approx 0.4\ t_{ri}$. This choice may be more cost effective because the extra dissipation during other parts of commutation may mask the increase caused by a slower diode. Note, however, as the diode lifetime becomes larger than the transistor's rise time, the power ratio increases rapidly.

Speeding the transistor rise time when using a slow diode may appear to improve matters, but it is not always wise. Assume that a circuit is operating with components such that $*T_S/t_{ri} = 2$. The power ratio is approximately 11. If the rise time is cut in half, the ideal power dissipation (the term outside the brackets in Equation 10.5) is cut in half, but the power ratio from Figure 164 is now 20. Clearly, little has been gained. In practice, much has been lost. The faster switching will produce a higher peak switch current because $I_{RM(REC)}$ will be higher; consequently, the power switch drive circuit will need to produce much higher peak currents and will consume more power.

In most designs, the power switch rise time is chosen to provide low loss at the frequency of operation. Losses caused by the diode recovery process are tolerable and a cost effective design is achieved if diode lifetime ($*T_S$) is chosen to be 20% to 50% of the transistor current rise time.

Practical Drive Circuit Problems

In the preceding idealized analysis circuit parasitic inductances were ignored. When switching at very fast rates, their effects cannot be ignored and may even dominate the switching process. During the t_a interval, dt/dt is determined by the transistor's characteristics, its drive and parasitic inductance common to the output current loop and the drive current loop. Voltage induced in the common inductance plays a significant role in switching at high speeds.

Consider the circuit of Figure 164. As current builds up in the common inductance, the developed positive voltage effectively reduces the drive voltage (negative feedback), placing a practical limit upon the ultimate switching speed during the t_a portion of diode recovery. At the end of time t_a, switch current drops to I_O. The di/dt during the t_b interval is not directly controlled by the transistor but is a function of the diode stored charge recovery characteristics as well as the diode's capacitance resonating with parasitic loop inductance. With ultrafast diodes, the recovery interval t_b is short and the transition abrupt. The resulting high di/dt causes a negative voltage across the common inductance which serves to increase the on–drive voltage. This induced voltage generates positive feedback and at high commutation speeds will greatly increase the diodes abruptness.

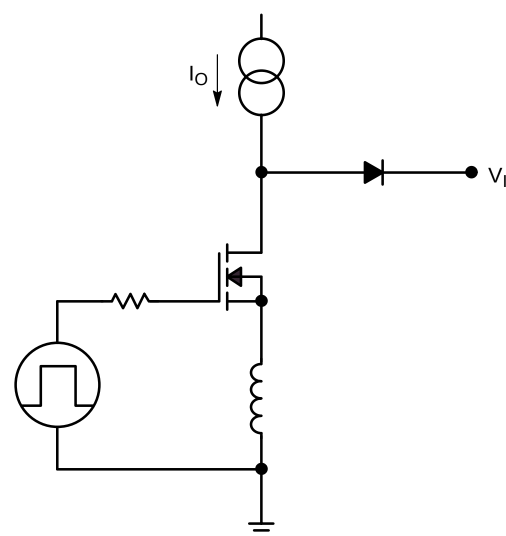


Figure 165. Inductance Common to Drive and Output Current Loops Slows *di/dt*

Another source of feedback is the capacitance from output to input of the power switch. When the voltage drops across the switch, a current is coupled into the switch input circuit. The coupled currents direction is opposite to that of the drive, which slows the voltage transition. The effect is

normally not noticeable when the switch is a bipolar transistor because the collector–base capacitance is small and the base current has a relatively high value; the rate of change of voltage is determined primarily by the components in the output circuit. However, the feedback capacitance of a MOSFET is large and, because of Miller Effect, constitutes the major portion of the FET's input capacitance. Because of the positive feedback noted above, a drive scheme selected for a suitably fast initial di/dt may cause the rectifier to snap off very abruptly. If the drive is reduced to slow di/dt to an acceptable speed, the drive may not provide a fast enough dv/dt to keep losses low during the t_b interval.

Thus, diode Q_{RR} and common inductance between the input and output current loops team up to compound the problems of obtaining a low–loss, low–noise turn–on of the power switch. Current and voltage snubbers, as discussed in the next chapter, can mitigate these effects. However, using a diode with low Q_{RR} and, if possible, a soft recovery reduces the magnitude of the problems.

Layout Considerations

High–speed forced commutation causes high di/dts and dv/dts in circuit wiring, which places stringent requirements on layout. A poor layout makes a high–speed rectifier look sluggish. For example, parasitic circuit inductance is often the underlying cause of what appears to be poor forward recovery characteristics. A high duck flowing through excessive stray inductance produces a large induced voltage. In a catch diode application where the purpose of the diode is to clamp voltage to a power rail, the inductance diminishes its effectiveness.

The layout contribution to parasitic inductance caused by wiring can be estimated from the length of the wiring loop. The self inductance of one of two parallel wires may be calculated from

$$L = (2\,1n\,d/r + 0.5)l \qquad (10.6)$$

where L = self–inductance of a cylindrical wire (nH)

d = distance between axis of wires

r = radius of wires

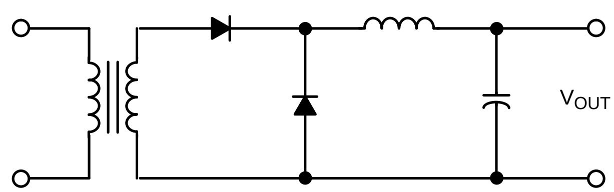
l = length of wires (cm)

The wires must be conducting equal current in opposite directions and d must be small compared to I for the equation to be valid. The minimum value occurs in the case of two enameled wires that are touching, making d = 2r. L calculates to be about 1.9 nH/cm. If d= l00r, the calculation yields approximately 10 nH/cm or 25 nH/inch, commonly used "rules of thumb." As the examples show, however, lower inductance is achieved by minimizing the distance between wires in a loop. If a transformer is in the circuit loop, its leakage inductance will be the dominant parasitic and is normally found by measurement.

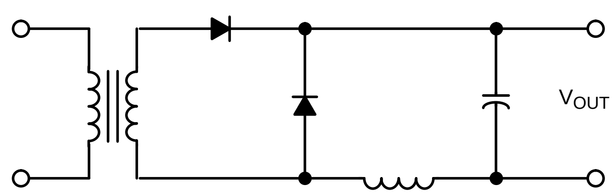
Another aspect of the problem caused by parasitic series inductance is a forced decrease in the maximum allowable commutation speed. As the transistor turns off in a hard switching circuit, such as a continuous conduction mode buck or boost converter, current in the diode increases because current in the output inductor is constant. If the di/dt and the parasitic inductance are high, the voltage across the power switch will be abnormally high. Therefore, if the transistor voltage is to be held in check, either its turn off must be slowed, which causes increased turn off losses, or a larger snubber is required, as discussed in the next chapter. Both of these remedies become increasingly difficult to implement as switching frequency is raised.

Parasitic capacitance across the diode is usually small relative to the rectifier's junction capacitance, so its effects are rarely seen. Capacitance from the diode's case to the heat sink must be addressed because the case is usually tied to the cathode.

Output rectifiers often experience high dv/dt, which injects noise into the heatsink when the output filter inductor lies between the rectifiers and the dc output as shown in Figure 166(a). By placing the choke in the return loop as shown in Figure 166(b), circuit operation remains unchanged since the choke is still in series with the parallel combination of the load and the output capacitors. The advantage of this rearrangement is that the cathodes of the rectifier diodes are not on a node that has rapid voltage swings and noise is not injected into the heatsink. The disadvantage of this arrangement is that the transformer return is now no longer at a dc node and radiation from the transformer must be addressed.



(a)



(b)

Figure 166. Output Rectifier Configurations:
(a) Conventional Arrangement for Minimizing Transformer Radiation;
(b) Alternate Arrangement for Low Noise injection into Heatsink

The Diode Snubber

Any of the forced commutation circuits usually produce high–frequency ringing in the circuit loop that has been switched off. The ringing occurs because the abrupt change of current excites the resonant tank composed on the effective circuit inductance (L_{EF}) and the diode junction capacitance (C_J). The Q of this network is high because loop resistance is purposely kept low to avoid power losses. Ringing may not appear in circuits using a diode which has soft recovery because its impedance while in the t_b phase damps the resonant circuit. However circuits using fast pn diodes with abrupt recovery or Schottky diodes usually exhibit severe ringing.

In circuits using Schottky diodes, the ringing energy can be high enough to generate a voltage higher than the diode breakdown voltage. Operation in breakdown is undesirable because of the increased power dissipation generated and the possibility of diode failure if its avalanche energy capability is exceeded. In addition, the EMI produced by the ringing usually causes malfunction in nearby electronic circuits.

To eliminate the ringing, the resonant circuit is critically damped by placing a series R–C network across the diode as shown in Figure 167. The capacitor serves to block dc and it is normally chosen to be about 10 times the capacitance of the diode (C_j). The resistor is chosen to provide critical damping; that is, the Q of the resonant tank is set to 0.5. For a parallel damping resistor (R_P) in the resonant tank:

$$R_P = Q\omega L_{EF} \qquad (10.7)$$

When L_{EF} is determined, the resonant frequency can be found from the usual relationship $\left(\omega = 1\sqrt{L_{EF}\,C_J}\right)$ and R_P determined from Equation 10.7. The capacitance C_J is voltage dependent, so that the ringing frequency is not a pure sine wave. A value of C_J picked from the diode's capacitance curve at about 1/3 of V_{RRM} usually provides suitable results.

Choice of the snubber capacitor (C_S) is a compromise between snubber effectiveness and snubber loss (P_R) that is dissipated in the resistor, R_P. The power loss is given by

$$P_R = 1/2\, C_S\, V^2_{RRM} f_S \qquad (10.8)$$

In order that the snubber capacitor is fully discharged each cycle, the time constant ($R_P\,C_S$) must be less than one–tenth of the minimum off–time of the diode. If R_P is made too low, it will feed significant signal around the diode. Therefore, efficiency is best when C_S is as small as possible and R_P as large as possible within the constraints given.

In a full–wave center–tap rectification scheme, one snubber across the secondary may be used when the output current is low. At high currents, the preferred method of one snubber across each diode is required to suppress the high–voltage spikes induced in the output wiring.

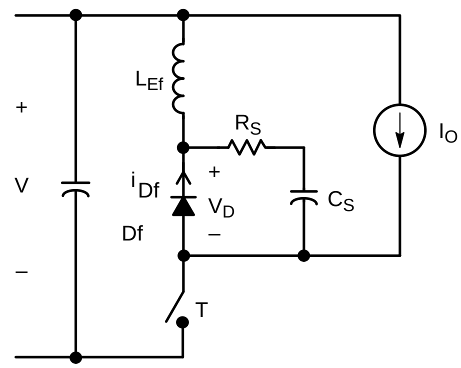


Figure 167. Equivalent Circuit Used to Analyze the Diode Snubber Circuit

The previous approach of using a large capacitor in the snubber network and choosing the resistor for critical damping prevents the voltage peak from exceeding the normal off–state level. Although satisfactory for low–voltage circuits, in higher–voltage circuits the power loss rn the snubber resistor is unacceptable. An approach which allows C_S to be reduced and shows the trade off in peak diode voltage has been devised [3] [4]. The peak voltage and resistor dissipation is obtained by analyzing the circuit of Figure 167, which is an equivalent circuit for nearly any hard commutation circuit.

A baseline capacitance C_b and a baseline resistance R_b are defined as

$$C_b = L_{EF}\left(I_{RM(REC)}/V_O\right)^2 \qquad (10.9)$$

$$R_b = V_O/I_{RM(REC)} \qquad (10.10)$$

The analysis neglects the diode junction capacitance and consequently the oscillations occur at a frequency determined by L_{EF} and D_S. The oscillatory wave is damped by R_S and the peak value is determined by R_S and C_S. For a given value of C_S, an optimum value of R_S can be found that minimizes peak voltage. Figure 167 shows behavior for $C_S = C_b$, which is ordinarily a good choice. Note that a broad null occurs when $R_S = R_b$, which yields a peak voltage of 1.5 V_O.

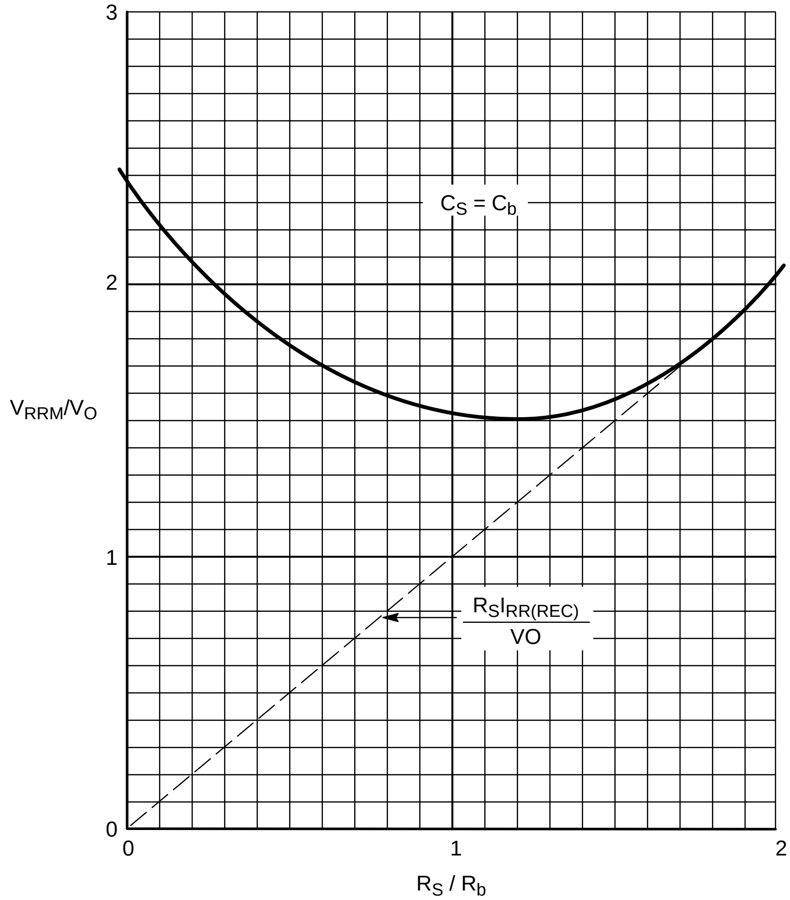


Figure 168. Maximum Overvoltage across the Diode as a Function of Snubber Resistance for a Fixed Value of Snubber Capacitance *CS*. (Courtesy Bill Robbins, University of Minnesota)

The energy loss in the resistor at diode turn–off consists of two components:

$$W_R = (1/2) L_{EF}\, I^2_{RM(REC)} + \left(1/2\, C_S V^2_O\right) \quad (10.11)$$

At the end of the current oscillations, energy (W_C) is stored in C_S, which must be dissipated in R_S and the diode when the diode turns on. This third component is:

$$W_C = (1/2)\, C_S\, V^2_O \quad (10.12)$$

The total energy dissipation W_T is the sum of the terms in Equations 10.11 and 10.12

$$W_T = W_R + W_C = (1/2)\, L_{EF}\, I^2_{RM(REC)} + C_S V^2_O$$
$$= (1/2)\, L_{EF} I^2_{RM(REC)} \left(1 + 2\, C_S/C_b\right) \quad (10.13)$$

Figure 169 shows the results of an extensive analysis of the circuit. All values are normalized, including the scales. Note that V_{RRM}/V_O decreases rather slowly from 1.5 to 1 as C_S becomes larger than C_b. However, the total energy dissipated increases linearly with C_S; therefore, it is common to choose $C_S = C_b$. When C_S has been selected, the optimum value for R_S is read directly from Figure 169.

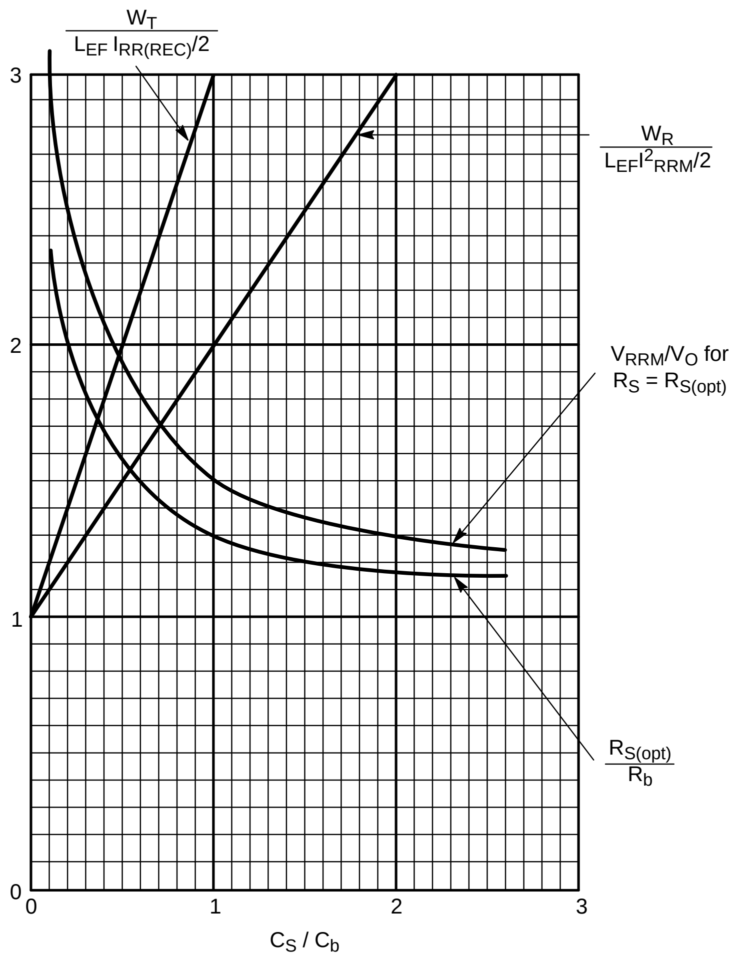


Figure 169. Snubber Energy Loss and the Maximum Overvoltage for the the Optimum Value of Snubber Resistance R_S and a Function of the Snubber Capacitance C_S (Courtesy Bill Robbins, University of Minnesota)

References

1. Rudolph P. Severans and Gordon E. Bloom, *Modern DC to DC Switchmode Power Converter Circuits*, Van Nostrand Reinhold Company, New York, N.Y.,
2. Ibid, 5, 6, & 7.
3. Ned Mohan, Tore M. Undeland, William P. Robbins, *Power Electronics, Converters, Applications, and Design*, John Wiley and Sons, New York, N.Y., pp. 470–478.
4. William McMurray, “Optimum Snubbers for Power Semiconductors,” IEEE *Transactions on Industry Applications*, Volume IA–8, No. 5, Sept/Oct 1972, pp. 593–600.

Chapter 11

Power Electronic Circuits

Power Electronic Circuits

In the previous chapter, several important functions that diodes fulfill are described. An analysis shows the advantages of using diodes with low stored charge, particularly in forced commutation circuits. In this chapter several circuits in common use are presented, emphasizing the role that diodes play and offering guidance in determining the characteristics of importance. Topics covered include drive circuits, snubbers, and rectification in switch–mode power supplies. By treating topics in the listed order, the interaction between circuit functions is more clearly perceived.

Drive Circuits for Power MOSFETs

In most cases, drive circuits for power MOSFETs are simple. A drive stage only needs to develop the required steady–state voltage between gate and source (usually 10 V) and provide sufficient peak current to the FET's input capacitance to achieve the desired switching speed. For noise immunity, the drive circuit should be a low impedance when the PET is in the off–state. The drive source return should be directly connected to the source terminal of the PET to minimize the effects of source inductance.

In some cases, electrical isolation is needed between the MOSFET and the drive stage. One way to do this is with a transformer. When the duty cycle of the drive signal varies over a wide range, dc restoration is needed to maintain proper voltage levels between the MOSFET gate and its source. A drive circuit with dc restoration is shown in Figure 170. In some circuits, a Zener diode may be required to prevent transient voltage spikes from damaging the MOSFET.

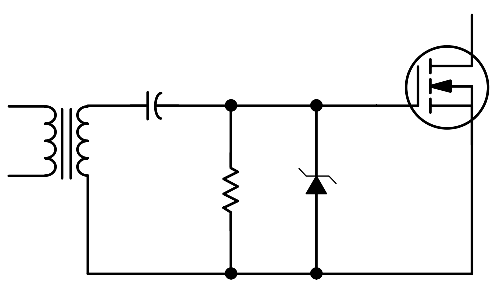


Figure 170. MOSFET Drive Circuit Using dc Restoration to Handle Large Changes in Pulse Duty Cycle

Drive Circuits for Bipolar Power Transistors

The design of a drive circuit for a bipolar junction transistor (BJT) operating in a power converter proves to be a challenging assignment. Criteria which generally apply are:

1. The on–drive (I_{BI}) should have a fast rise time and provide adequate drive at the leading edge to reduce turn–on losses.
2. After the initial peak, the amount of base current should drop to a level just sufficient to produce a satisfactory on–state ($V_{CE(SAT)}$) voltage. Keeping I_{BI} low during the major portion of on–time helps reduce the storage time. Since the collector current varies over a wide range in most converters, a proportional drive scheme or Baker Clamp helps achieve this condition.
3. A negative base drive (I_{B2}) is mandatory in order to keep turn–off switching losses to an acceptable value. The drive needs to be tailored to the needs of the particular transistor type used for the power switch. Most third and fourth generation BJTs perform well if the change from I_{BI} to I_{B2} is rapid and I_{B2} is high throughout the turn–off interval.
4. When the power transistor is off, the drive circuitry should present a low impedance to the transistor base in order to have good noise immunity.

Rectifier diodes play important roles in many of the commonly used drive schemes. It is easy to generate the required negative base current if a bias supply having a reverse polarity with respect to the emitter of the power switch is available. However, such a bias supply is seldom available, especially in a switch–mode power supply. Designers are loathe to create one because it complicates the design, particularly if the supply is to operate on worldwide powerline voltages. Consequently, the circuits presented in this chapter provide the means to generate the reverse base current without an external supply.

Baker Clamp

The storage time associated with the turn–off of bipolar transistors has always been a problem for circuit designers. Allowance must be made for worst–case storage time or disastrous events may occur. Such allowance lowers efficiency because it shortens the time available to deliver power pulses, which in turn raises the peak current. Unfortunately, storage time varies considerably with the values of I_C and I_{BI} prior to turn–off and with the amplitude and waveshape of I_{B2}. A technique developed many years ago by R.H. Baker is shown in Figure 171. It reduces and stabilizes storage time as operating conditions vary.

Circuit operation is as follows. When a positive pulse is applied to an initially off transistor, forward current flows through D_B into the base of the transistor, turning it on. The voltage at the input is set at two diode drops, approximately 1.5 V. As collector current builds up and collector voltage drops, diode D_F starts to conduct and keeps V_{CE} from falling below about 0.8 V. This action prevents the bipolar operating point from entering the hard saturation region and theoretically eliminates storage time since the collector–base junction is not reverse biased. Diode D_F acts as a current–regulating feedback element, setting the base current to the amount required by the transistor at a

particular collector current and on–state voltage. Current at the input which exceeds this base current (i.e., excess current) is shunted into the collector circuit by diode D_F. When the input pulse goes negative, reverse current is drawn from the base through diode D_R in parallel with diode D_B until its stored charge is depleted. Resistor R is present mainly for noise immunity; it also establishes a small reverse off–state voltage at the base if the input level remains negative.

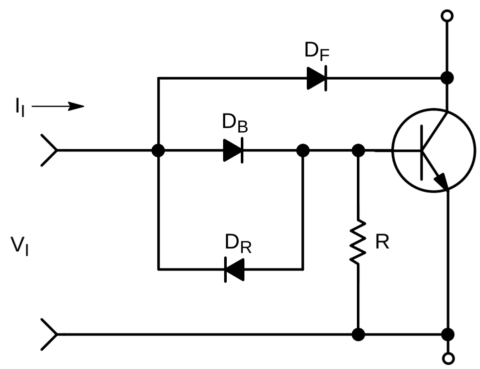


Figure 171. Baker Clamp Configuration

The Baker Clamp provides the expected results for low–voltage bipolar transistors – that is, storage time becomes negligible and fall time is not affected. However, high voltage power transistors exhibit a quasi–saturation region caused by stored charge in their wide, high resistivity collector region. When operated in quasi–saturation, some storage time still occurs. It is possible to move operation out of the quasi–saturation region by stacking additional diodes in series with D_B, but the on–state loss becomes unacceptably high.

It might seem that the Baker Clamp is not particularly effective with high–voltage transistors; however, operation in hard saturation causes anomalous fall time behavior, while fall time behavior is more theoretical if operation extends only into the quasi–saturation region. To show the effectiveness of a Baker Clamp [1] the test circuit of Figure 172(a) was used. Transistors Q_1 and Q_2 are alternately turned on to provide I_{B1} and I_{B2} to the power switch. The load inductor, diode, and capacitor simulate a flyback converter load. A sketch of the waveforms observed are shown in Figure 172(b). When I_I is first applied, it is delivered solely to the transistor base. Current builds up slowly in the inductor but its high impedance allows collector voltage to drop quickly. Diode D_F conducts, shunting the excess base current into the collector of the power switch. As collector current (I_{CT}) rises, base current (I_{B1}) drops and maintains the value I_C/h_{FE} as collector current ramps toward its peak, which is established by the pulse width of I_I.

The most interesting results occur with a fixed base drive and a variable collector current, a common situation in a switching power supply. The effect upon storage time is shown in Figure 173. Note that use of the Baker Clamp has achieved a remarkable reduction in storage time at low collector currents and its value is somewhat independent of collector current. Improvement in fall time is even more spectacular as shown in Figure 174. The trade–off made for the improved switching performance is an increased on–state voltage. It is typically 0.3 V in hard saturation and 0.9 V with the Baker clamp.

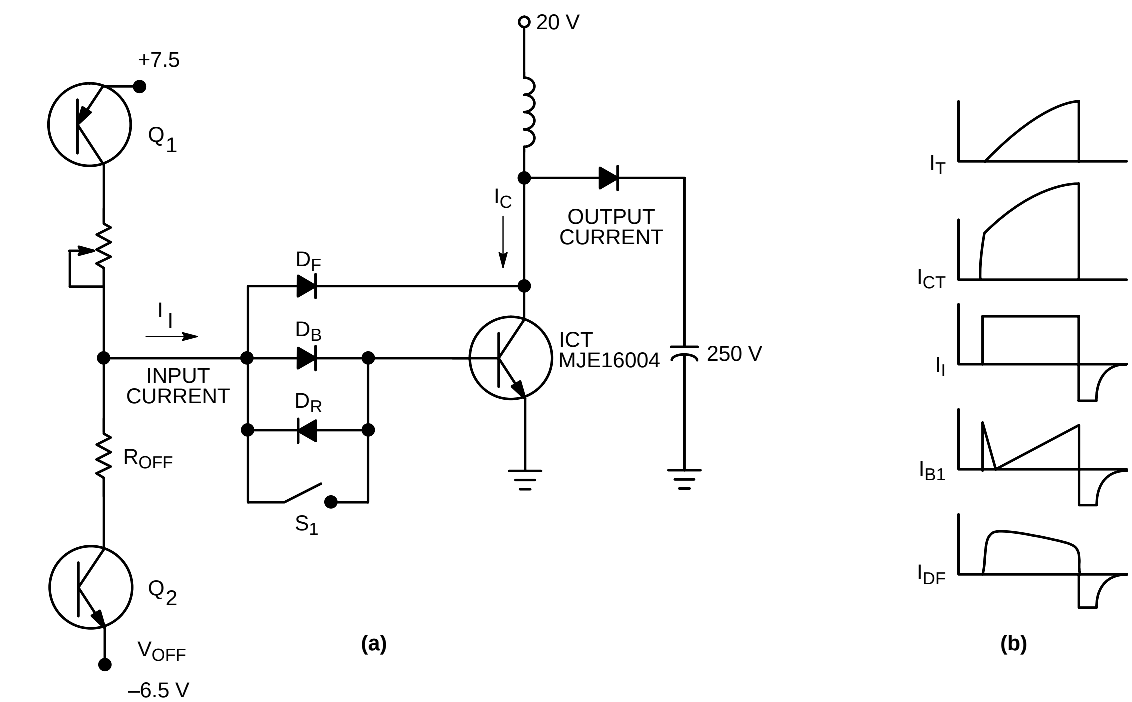


Figure 172. Baker Clamp Circuit for Evaluating High–Voltage Transistors: (a) Test Circuit; (b) Waveforms

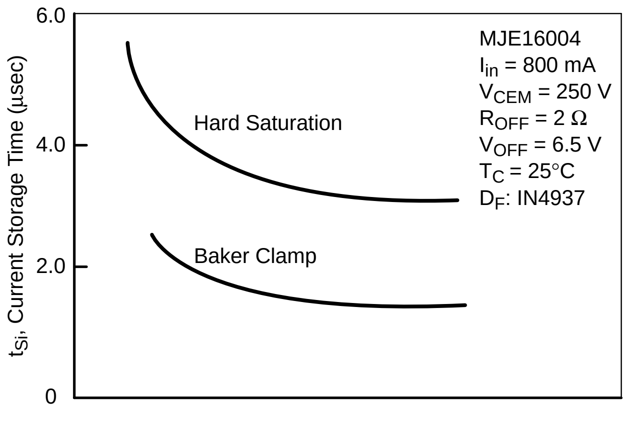


Figure 173. Storage Time with Fixed Drive in a Baker Clamp Circuit

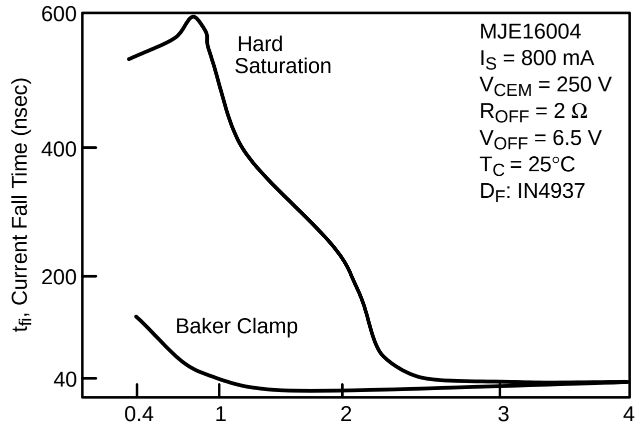


Figure 174. Fall Time with Fixed Drive in a Baker Clamp Circuit

Not all transistors will exhibit the behavior illustrated in Figures 173 and 174, and data of this nature rarely appears on manufacturer's data sheets. Consequently, it is advisable to test transistors in a circuit similar to Figure 172 to determine if the improvement in switching losses more than offsets the additional on–state losses associated with the Baker Clamp.

Selection of the diodes deserves critical attention. The collector feedback diode D_F needs to be able to block the voltage applied to the collector of the transistor. Its stored charge should be low enough that the diode recovers during the storage time of the transistor. Otherwise, it will play a role in the turn–off process, affecting both storage and fall time. Finding the right diode may not be easy since stored charge is not often specified and conditions for reverse recovery specifications do not approximate those in a Baker Clamp circuit. As a rough rule of thumb, the specified reverse recovery time should be under a quarter of the specified storage time of the transistor. Testing a number of samples with a current probe in series with diode D_F will normally be necessary to be sure that an adequate margin exists between diode recovery time and transistor storage time. A slow forward recovery time is actually beneficial. The delay allows the spike of current from I_I to persist longer, thus reducing turn–on loss. Fortunately, slow forward recovery is a characteristic of high–voltage diodes, regardless of their stored–charge behavior.

On the other hand, fast forward recovery is important for diode D_B, unless a peaking capacitor is placed in parallel with it. A low–voltage diode provides fast forward recovery. A large area diode also may provide sufficient peaking capacitance. Slow reverse recovery is an advantage. The stored charge will hold the anode positive during the turn–off phase, thereby increasing the current available from the negative voltage source. Often diode D_R is omitted; the circuit then depends upon the stored charge in diode D_B to exceed the stored charge in the transistor, thereby providing a current path for I_{B2} until the transistor turns off. Diode D_R also has non–critical requirements. It is used simply to apply reverse current (I_{B2}) when V_I is derived from a scheme which drives V_I negative.

Fixed Drive Circuit

The circuit of Figure 175 is commonly used with circuits such as the half bridge in which the power switch cannot be connected to "ground" [2]. It generates a negative pulse without the need for a separate supply voltage. Furthermore, the rise time of the reverse current is not limited by transformer leakage inductance.

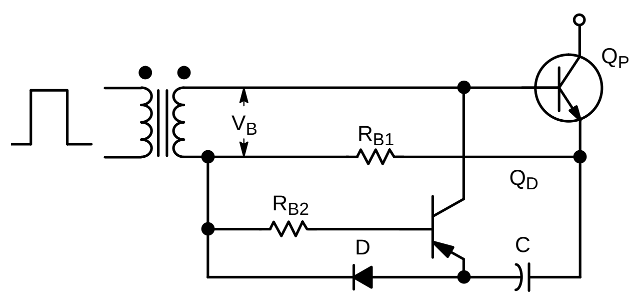


Figure 175. Popular Fixed Base Drive Circuit Using an Isolation Transformer

The circuit operates as follows. When a positive pulse is applied, a high peak forward current flows through the capacitor and diode into the base of Q_P, turning it on. As C becomes charged, I_{B1} drops to the value determined by resistor R_{BI}. I_{B1} is given by:

$$I_{B1} = (V_S - V_{BE})/R_{B1} \qquad (11.1)$$

and the voltage on C is:

$$V_C = V_S - V_{BE} - V_D \qquad (11.2)$$

where V_S = amplitude of transformer secondary pulse
V_{BE} = base–emitter voltage of transistor Q_P
V_D= forward voltage drop of diode D.

When the pulse voltage drops to zero, the voltage on capacitor C forward biases the driver transistor, Q_D. As Q_D turns on, the negative voltage on C causes a high reverse current (I_{B2}) to flow out of the base of Q_P.

The circuit can generate a fast–rising, high–current I_{B2} pulse. It is important that parasitic wiring inductance be kept to a minimum. The drive transistor Q_D should be a high–speed switching type and capacitor C must have a low ESL and ESR. Diode D does not affect the turn–off drive, but it should have a fast forward recovery time and low V_{FRM} in order to apply a high I_{B1} to the transistor during turn–on. Care must be taken to ensure that C is fully charged during the minimum pulse width, which the transistor is expected to handle.

Proportional Drive Circuit

Another method of minimizing the increase of storage time with a decrease of collector current involves arranging the base drive current to be proportional to the collector current. Not all bipolar transistors respond favorably with proportional drive, but most do. In addition, high–voltage transistors usually exhibit an improvement in fall time, which lowers switching losses. The improvement in turn–off time is similar to that achieved with the Baker Clamp, but the transistor remains in the "hard" saturation region of operation when in the on–state, which achieves lowest on–state losses.

The circuit of Figure 176 has been used [3] to drive bipolar transistors in a power converter operating up to 1 kW of output power. The scheme is often called a regenerative drive circuit because the secondary windings of the drive transformer are in phase, allowing the power transistor to supply its drive signal. During steady–state conditions when the power transistor Q_P is on, the driver transistor Q_D is off, causing the primary winding to be inactive. The base current is derived from the collector current and its magnitude is determined by the turns ratio of the secondary windings. Consequently, no drive circuit power is required to hold the power transistor on, which results in an efficient drive scheme. The Baker Clamp diode D_F may or may not be used to improve overall losses, depending upon circumstances.

The circuit operates as follow: Assume the power switch Q_P is on and Q_D is off.

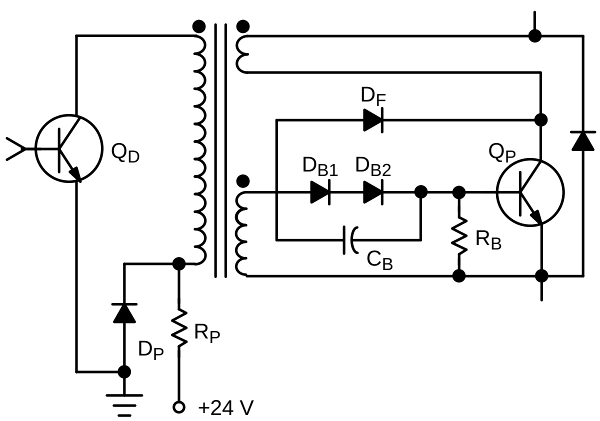


Figure 176. Efficient Proportional Base Drive Circuit Which Easily Adapts to Using a Baker Clamp

The primary winding has a large voltage drop across it caused by the current flow in the secondary base drive winding which produces a drop of about 2.5 V across the series diodes and the base–emitter junction of Q_P. In addition, C_B is charged to about 1.6 V. The phasing of the windings is such that the dotted ends are positive. When Q_D is turned on into hard saturation, the dotted end of the primary is pulled close to ground. The other end of the primary winding would go negative but cannot because of the presence of clamp diode D_P which turns on and clamps the winding to ground. The net effect is one of a short being placed across the primary which stops the regenerative action in the secondary. The secondary windings now appear as a short. Therefore, a reverse base current is supplied to the power switch as C_I discharges and the stored charge in diodes D_{BI} and D_{B2} exits. During turn–off of Q_P, primary current flow is composed of the reflected currents in the secondary windings which flow through D_P and Q_D. The base drive to Q_D must be sufficient to keep it in hard saturation while handling the relatively high reflected secondary currents.

After Q_P is off, the drive transformer primary current decays to zero and diode D_P turns off. Current flows through R_P and the primary winding which places a reverse bias on Q_P, holding it off. Since the only current flow is in the primary, energy is stored in the core. When the system requires Q_P to turn on, Q_D is turned off and the stored energy in the core develops a flyback voltage of a polarity to turn on the power switch, Q_P. The resulting build–up of collector current causes a proportional regenerative build–up of base–drive current (I_{B1}). which maintains the power switch in the on–state after the flyback energy has dissipated.

Switching performance is highly dependent upon the choice of components in the drive circuit as well as the choice of the power switch. The transformer leakage inductance must be small to permit a rapid change of base drive from I_{BI} to 1_{B2}. The windings are usually trifilar wound on a toroidal core. Naturally, leads and wiring traces must be short to minimize inductance. The return lead to the base winding should be connected directly to the emitter of the power sWitch to prevent the relatively large voltage developed across the emitter lead inductance from being introduced into the base circuit. Diode D_P should have a low turn–on impedance since its reflected impedance is part of the base circuit during the turn–off interval of the power switch. A low–voltage pn or Schottky diode serves very well in this application.

Diodes D_{B1} and D_{B2} are basically level shifters so their switching characteristics are unimportant. Their function, however, is very important because they established the voltage across the drive transformer windings when the power switch is on. A resistor cannot take their place because its drop would vary with changes in load current.

The circuit easily adapts to Baker Clamp operation by merely adding diode D_F. The improvement in switching performance is not as dramatic as achieved in fixed base drive circuits; with many transistors, lower overall losses are achieved by letting the transistor saturate. The higher the switching frequency, the more likely is the Baker Clamp to be an effective scheme to lower losses. If the Baker Clamp is used, diode D_F should be selected for low stored charge, as previously discussed.

Transistor Snubber Circuits

Networks are frequently added to power electronic circuits to control the load line and thereby reduce the energy which a power switch must handle during switching. These networks are called snubbers, and diodes usually play a key role. Snubbers are commonly classified into three types based upon their function:

1. Turn–on snubbers
2. Turn–off snubbers
3. Overvoltage snubbers

Additionally, these functions can be accomplished by using either dissipative or nondissipative techniques for handling snubber energy. In a dissipative scheme, energy that would cause transistor dissipation is, in effect, transferred from the power switch to a resistor. In a nondissipative or “lossless scheme,” resistors are avoided and transistor switching losses are transferred to the source or the load using some combination of diodes, capacitors inductors or transformers.

Power Circuit Load Lines

The usual load line of a power circuit is inductive. Furthermore, a diode typically is added to catch inductive spikes and it may also perform the freewheeling function when operation is in the continuous conduction mode. To the transistor, a load of this nature is referred to as a clamped inductive load.

The buck converter previously discussed in Chapter 10 shows the need for snubbers. An equivalent circuit is shown in Figure 177, where the filter inductor is replaced by a current source (continuous mode operation) and stray inductances are shown as separate circuit elements. Transistor switch waveforms for a complete switching cycle commencing with the transistor in the on–state are shown in Figure 178.

At time t_0, the transistor voltage begins to rise but its current remains essentially constant. At t_1, the freewheeling diode starts conducting, which diverts current away from the transistor. The voltage across the transistor overshoots V_I because of diode forward recovery time and circuit inductance. In a high–voltage circuit, transistor turn–on is faster than diode forward recovery time so that i_c reaches zero at t_2 and the transistor collector–emitter voltage settles to its steady state value at t_3. The turn–off switching path on the V–I plane is seen from Figure 179 to cause the transistor to experience high instantaneous power dissipation. Depending upon the switching speed and frequency, the average power dissipation could also be significant. For bipolar transistors, the high power dissipation may be outside of the transistor’s safe operating area, posing danger of failure. The bipolar transistor is particularly sensitive to voltage spikes during turn–off; thus the stray inductive spikes and the forward recovery overshoot voltage of the freewheeling diode (V_{OS}) critical to transistor survival.

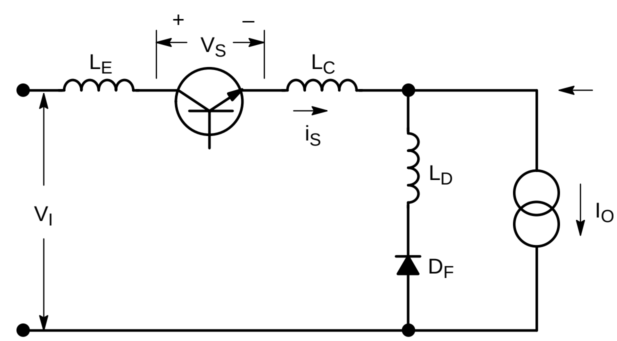


Figure 177. Equivalent Circuit of a Buck Converter Operating in the Continuous Conduction Mode

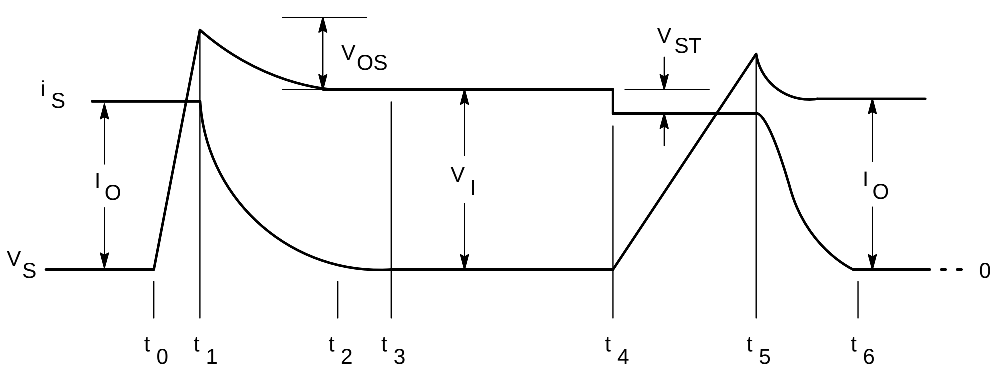


Figure 178. Power Switch Waveforms for the Buck Converter

The transistor's turn–on period commences at time t_4. As current rises, the stray inductances generate a voltage step, V_{ST}. As described previously in some detail, the reverse recovery current of the diode causes the switch current i_s to exceed I_O. At t_5 the diode impedance begins to increase which allows the collector voltage to drop towards the on–state level, which it reaches at time t_6. The turn–on path shown on Figure 179 is also seen to produce high instantaneous dissipation. The situation is not as stressful to the power switch because semiconductors are more forgiving of current peaks than voltage peaks.

Snubbers reduce these power peaks by pushing the switching paths toward the origin. If the paths shown on Figure 179 protrude beyond the safe operating area of the power switch, a snubber is essential in order to insure reliable operation. Even if not required to protect the power switch, snubbers are often used to reduce switching losses in order to allow the power switch to run cooler or lessen heat sink requirements. Snubbers, when carefully chosen, result in a higher power efficiency for the system than if they are omitted, even if the dissipative type is used [4].

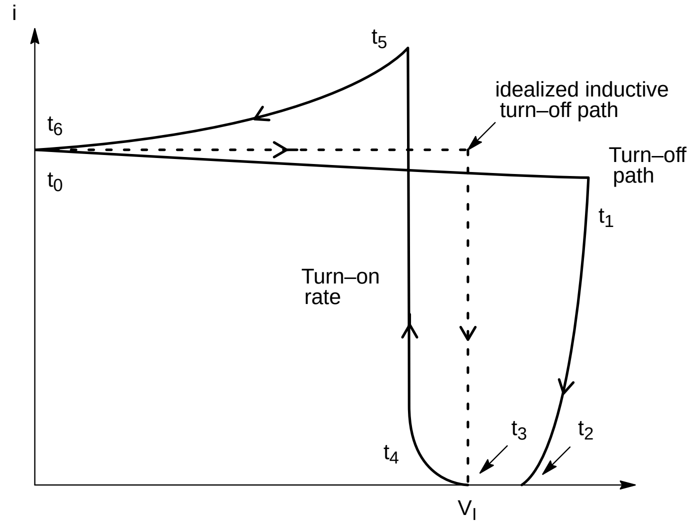


Figure 179. Switching Load Loci for the Buck Converter

Turn–On Snubbers

Turn–on snubbers are used more frequently with thyristors than transistors because thyristors have a finite turn–on di/dt limitation, whereas transistors have large turn–on (forward biased) safe operating areas. However, the turn–on snubber lowers stress and peak power dissipation in the power switch. It is particularly effective with high–voltage bipolar transistors and thyristors which exhibit a significant dynamic saturation voltage. The turn–on snubber makes use of an inductor to slow current rise time and allow a rapid drop in voltage across the switch to occur prior to current build–up. Figure 180 shows how a turn–on snubber is implemented in the buck converter. The key requirement is that the inductor is placed in the current path between the power switch and the freewheeling diode.

Two modes of operation are possible. The first mode occurs with a small inductance such that $\Delta V_S = L di_S/dt < V_I$. The second mode occurs when $\Delta V = V_I$.

In the first mode, di_s/dt is determined by the characteristics of the power switch and its drive circuit. The voltage step, V_{ST} on Figure 178, becomes progressively larger as L_S is increased, but the current waveforms of the transistor and freewheeling diode are virtually unchanged. The turn–on path from t_4 to t_5 is moved toward the left on Figure 179 as L_S increases.

In the second mode, di/dt is controlled by the inductor L_S. The current switching time is relatively slow but switch dissipation is nearly zero because the voltage drops almost immediately to the off–state level as current starts to rise. With adequate drive to the switch, dynamic saturation is not noticeable.

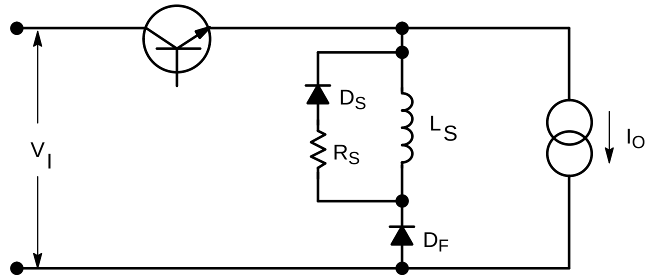


Figure 180. Buck Converter with Dissipative Turn–on Snubber

The current overshoot caused by the reverse recovery of the freewheeling diode can be reduced to any value desired by the choice of L_S. The turn–on path from t_4 to t_5 is a vertical line originating close to the origin. When the transistor is turned off, the energy stored in L_S during the switch on–state must be controlled. The diode D_S is used as a catch diode to clamp the inductive spike generated at switch turn–off. Unfortunately, the diode cannot be used without a series (R_S). The reason is that the inductor current must decay nearly to zero during the switch off–state in order for the snubber to be effective. Accordingly, the snubber circuit values are bound by the following inequality:

$$t_{(off\ -\ state)} > 3\ L_S/R_S \qquad (11.3)$$

The resistance of the diode alone is ordinarily too low to allow proper circuit operation at the high frequencies used in power electronic circuits. Unfortunately, an overshoot voltage equal to $R_S I_O$ is developed across the resistor R_S. Thus, the design of the turn–on snubber involves a number of compromises. The snubber's maximum time constant (L_S/R_S) is limited by the frequency of operation. Making L_S larger results in lower turn–on losses, but since R_S must also become larger, the overshoot will increase proportionally.

The inductor used for L_S is not physically large because it is allowed to saturate at t_5 since its function is required only during the interval from t_4 to t_5. The diode D_S performs a catch function; therefore its forward recovery overshoot voltage is important because it adds to the overshoot voltage developed across the resistor R_S.

A nondissipative or "lossless" turn–on snubber is occasionally used. One approach, based on the work of Meares [5], is shown in Figure 181. The snubber inductance is the transformer's magnetizing inductance which performs the same functions as the discrete inductor in Figure 180 and is subject to the same design constraints. When the current reverses through the freewheeling diode, the current is coupled through the transformer whose secondary delivers the reverse recovery current to the input through the snubber diode D_S. In this manner, the current demand from the power source is reduced.

The scheme is not truly "lossless" because losses in the transformer and diode D_S cannot be avoided. The forward recovery of D_S is more important than its forward drop because it only conducts current for a short period of time. Its specification requirements are similar to that of a catch diode.

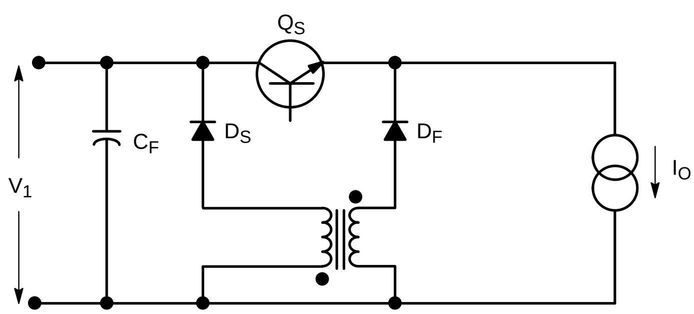


Figure 181. Buck Converter with "Lossless" Turn–on Snubber

Turn–Off Snubbers

The turn–off snubber is designed to slow the rise in voltage across the power switch while allowing the current to drop rapidly. The most common approach uses an R–C–D network as shown in Figure 182. When the power switch is on, its current equals I_O and the voltage drop across the switch is nearly zero. Since I_O is constant during turn–off, as collector current drops, a current $I_O–i_C$ flows into the capacitor through the snubber diode D_S. The capacitor is chosen to achieve the desired rate of rise of the transistor collector–emitter voltage. The resistor is used to charge the capacitor when the switch is on, and therefore it must be bound by the following inequality:

$$t_{(on\,-\,state)} > 3\,R_S/C_S \qquad (11.4)$$

To arrive at a suitable design, the peak current to discharge the capacitor must be considered. It may be approximated simply as $V_{CE(off)}/R_S$. Thus, for a given frequency of operation the R_SC_S product has an upper bound. As C_S is made larger to reduce the turn–off energy consumed in the power switch, R_S must be smaller which increases the turn–on energy. For a given set of conditions, an optimum value of C_S exists which minimizes the total losses in the power switch and snubber over a complete cycle [4]. Using the optimum value makes the load line appear resistive.

The snubber diode's forward recovery characteristic is of major interest since the diode appears as a time variable resistance (r_F) in series with the capacitor during power switch turn–off. Therefore, in addition to the capacitor voltage, a voltage of $(I_O–i_C)\,r_F$ appears across the transistor, increasing its dissipation.

Turn–off snubbers can also be designed to be "lossless." They make use of a resonant circuit; the complete analysis [6] is complex and beyond the scope of this treatment. However, a qualitative description of a popular implementation shows the importance of diode characteristics, which is the focus of this chapter. A practical circuit arrangement is shown in Figure 183(a) where the lossless snubber is applied to a flyback converter. The snubber could be applied as well to other transformer coupled topologies.

The snubber components are D_{S1}, D_{S2}, L_S, and C_S. The transformer leakage inductance reflected to the primary is shown as an external component, L_L. The turns ratio of the transformer is n:1 and its magnetizing inductance is very large compared to its leakage inductance, so that the magnetizing inductance does not affect snubber operation.

A qualitative view of basic circuit operation is easily obtained by assuming that the power switch Q_S is initially in the on–state. The capacitor is charged to some positive voltage between zero and V_I. When the drive signal changes to initiate turn–off, after a short delay switch current starts to fall. The inductive load keeps the load current constant causing current to switch from Q_S into C_S and D_{S1} since the current in L_S cannot change rapidly. Depending upon the initial charge on C_S, the switch voltage may exhibit an initial jump until diode D_{S1} becomes forward biased. Once D_{S1} is conducting, the rate of rise of voltage across the switch is controlled by C_S in a manner similar to that of the more conventional R–C–D snubber. With a properly chosen capacitor having a sufficient initial charge, the switch voltage rises only to a chosen small portion of the off–state voltage before the switch current reaches zero. The forward recovery characteristics of the diode are important since its transient voltage is reflected to the switch voltage waveform.

When Q_S is turned on, it completes the series resonant circuit composed of Q_S, C_S, D_{S2}, and L_S. The charge on C_S and its voltage polarity is such that current flows through the network, and the voltage at point A drops to $-V_{CP}$, the voltage across the capacitor prior to the turn–on of Q_S. As current rises in the resonant loop, the voltage at point A rises toward $+V_{CP}$ but D_{S1} conducts when V_A exceeds V_I. At that time, the energy stored in L_S flows into the input source. Thus, the energy stored in the capacitor during switch turn–off is transferred to the inductance and then to the input source when the power switch again turns on.

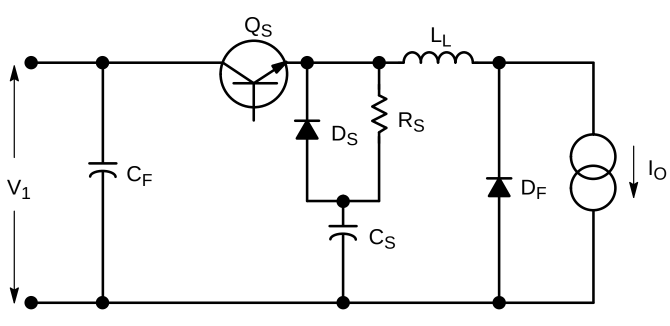


Figure 182. Buck Converter with Dissipative Turn–off Snubber

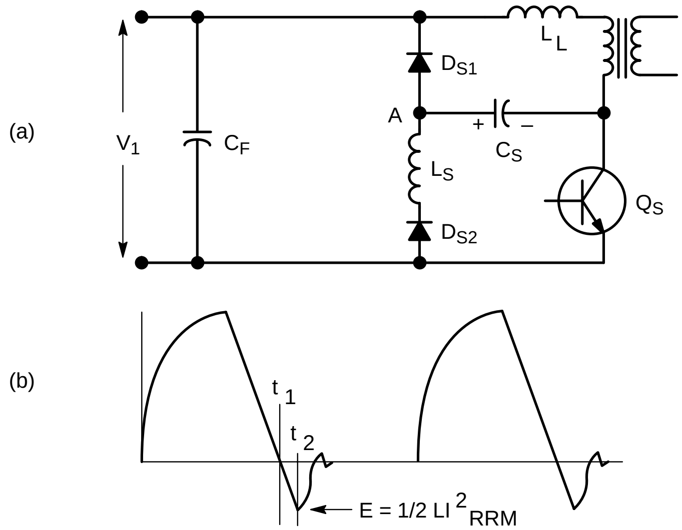


Figure 183. Application of a "Lossless" Snubber of the Resonant Type: (a) circuit; (b) current in D_{S2}

The mode of operation described is based upon one of the earlier papers on the subject [7]. A later paper [6] explains that four different modes of operation exist depending upon the choice of snubber constants, the transformer design and frequency of operation. While differing in details, all modes allow proper operation of a power converter if suitably designed.

Resonant snubbers are difficult to implement in converters operating under current mode control because a significant amplitude of half–sine wave current at the snubber resonant frequency appears on the power switch waveform during turn–on. The problem must be addressed or converter operation will be erratic.

In descriptions of circuit operation, diode requirements are seldom explained in detail. For the circuit being considered, abrupt current changes only occur when the power switch changes state or when the diode turns on. Consequently, a fast diode turn–on characteristic reduces losses. Diode reverse recovery is usually important in D_{S2} but not D_{S1}. The current waveform of D_{S2} is shown in Figure 183(b). The turnoff di/dt is controlled by L_S. Normally the di/dt is high enough that significant reverse recovery stress occurs. The diode lifetime (*T_S) should be at least three times less than the resonant period of the snubber to avoid significant diode reverse recovery losses.

Overvoltage Snubbers

In most power electronic circuits, even when the load is clamped, parasitic inductances cause overvoltage spikes. These parasitics are introduced because of transformer leakage inductance, capacitor effective series inductance (ESL), and wiring inductance. A turn–off snubber does not always eliminate the overvoltage spike because of the parasitic inductances in the snubber circuit. Furthermore, many circuits have no need of a turn–off snubber because of the load characteristics in relation to the power switch SOA and because operation is at a frequency where the switching power loss is not important. In these cases, an overvoltage snubber may prove useful.

Overvoltage spikes can be clipped with the circuit of Figure 184. In operation, while the power switch is off, the voltage on the overvoltage capacitor C_V is pulled up to the supply voltage, V_I, by diode D_V. When the switch turns off, the diode D_V is activated whenever the energy in the small leakage and stray inductance, lumped as L_L, causes the switch voltage to exceed V_I. The switch voltage is clipped to V_I plus the diode forward recovery voltage plus the voltage increase on the capacitor as it absorbs the inductive energy. The time constant R_VC_V is chosen small enough to allow the capacitor voltage to decay back to the normal off–state level before the next turn–off cycle of the power switch.

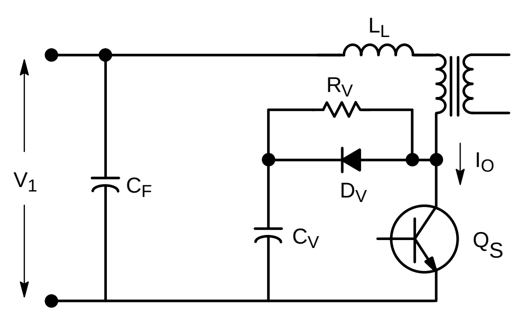


Figure 184. An Overvoltage Snubber

With some loads, the switch voltage during the off–state will exceed V_I under normal operation. For instance, a forward converter is typically designed to allow an off–state voltage of $2V_I$. The overvoltage snubber will work in many of these situations because after a few cycles of circuit operation, the capacitor C_V will be charged to the normal

off–state voltage. Any short spikes which exceed this level are clipped.

While clipping, the energy from the net circuit inductance is delivered to the capacitor (C_V). The voltage increase across the switch (ΔV) is found by equating energy to obtain:

$$\Delta V = I_0 \sqrt{L_L / C_V}$$

To calculate the switch voltage, the forward recovery voltage of diode D_V is added to ΔV from the equation. To minimize the spike, it is very important to keep inductance low in the loop consisting of Q_S, C_V, and D_V. Use of a high–quality capacitor having a low ESL is beneficial.

An alternate method of clipping spikes is to use a transient voltage–suppression type of Zener diode. (The MOV, discussed in Chapter 9, has too wide a voltage tolerance and too high a dynamic impedance to function well in power electronic circuits.) However, the Zener may dissipate more energy than the resistor because it must operate at a voltage level greater than V_I. Using the overvoltage snubber, the energy is dissipated in R_V. It only needs to handle the change in voltage as given by the above equation.

Snubber Diode Requirements

From the descriptions of snubber diode requirements presented in the circuit examples, it is clear that the diode's forward recovery behavior is of most importance in snubber operation. Specifically, a low dynamic resistance is required at turn–on. Unfortunately, forward recovery is not a required specification on JEDEC registered devices and it is rarely specified on manufacturer's data sheets.

As explained in Chapter 1, the diode dynamic resistance r_F at turn–on is determined by the resistivity of the materials in the diode structure and its area. The breakdown voltage is roughly proportional to resistivity. Therefore, the turn–on performance of diodes improves as voltage breakdown decreases. Little improvement is seen as breakdown drops below 100 V, however, because other sources of bulk diode resistance begin to dominate.

Diode selection for low dynamic turn–on resistance on the basis of voltage break–down is not straightforward. The specified voltage rating is not a reliable guide to the actual breakdown voltage of a rectifier diode because of a long–standing industry practice of sourcing both high and low–voltage devices from the same product family.

The exception is the Schottky diode, which has a low–voltage rating because of the difficulty in making high–voltage parts. Consequently, the actual breakdown voltage of a Schottky is only slightly in excess of rated voltage. Using Schottkys for applications requiring voltage ratings below 100 V yields best performance. For higher voltage applications, the choice is less obvious. Some ultrafast diode families have limited voltage capability which might provide low dynamic impedance. To be safe, the manufacturer should be consulted. Other sources for low dynamic impedance diodes having breakdown voltages up to 200 V include zener or transient voltage suppressor diodes. These diodes have well controlled voltage breakdown levels but are not processed for fast reverse recovery and therefore cannot operate at as high a frequency as an ultrafast device. From 400 to 1000 V, tests [8] have shown that a MEGAHERTZ™ rectifier provides improved performance.

Power Conversion Circuits

The output rectification section of a switch mode power supply (SMPS) commonly uses a full–wave single–phase circuit (as described in Chapter 5) only when the input is an alternating wave – that is, a wave having both positive and negative half cycles of equal amplitude and duration. These waves are produced by various kinds of push–pull power converter arrangements. However, a number of popular topologies are single–ended; therefore, the rectifier circuit is fed with unidirectional power pulses. In these cases, a half–wave rectifier is satisfactory, and means are employed to reset the transformer core to prevent saturation.

The major use of SMPSs is to power digital equipment whose logic circuits commonly operate from a 5 V bus. Linear circuits and disk drives usually require other voltages, with 12, 15, 24, and 28 V being in common use. SMPSs are cost effective at power levels as low as 25 W and are also used to provide power up to several thousand watts. Obtaining efficient high–frequency rectification at high power and low voltage is a design challenge. For example, the rectifiers in a 5 V, 1 kW supply must handle 200 A, which dictates use of low forward drop Schottky diodes. Additionally the equivalent series resistance (ESR) of capacitors usually determines ripple which forces use of capacitors of much larger value than indicated by the design curves in Chapter 7. Attention must also be given to bussing and connections in order to minimize losses.

In this section, rectifier circuits commonly used with popular topologies are discussed with the purpose of understanding rectifier diode requirements and penalties imposed by using diodes having various types of nonideal behavior. In addition, other information helpful in the design of high–frequency rectifier circuits is offered. This treatment assumes the reader is familiar with operation of the various topologies commonly used in switch–mode power supplies and therefore descriptions and equations are presented without qualification. Necessary background is provided by the books listed in the bibliography.

The following material shows how rectifier diodes are matched to the requirements of some popular converter topologies used in switching power supplies. The treatment is not exhaustive to keep it within the scope and space allocated within this book. However, the issues involved in diode selection are felt to be sufficiently discussed so that the principles used can be applied to other topologies in addition to the ones shown.

Flyback Converter

The output section for the flyback converter in Figure 185 shows the power switch Q_S, snubber network (R_S, C_S, and D_S), the output rectifier diode D_R, and filter capacitor D_F. The circuit can be operated in either the continuous or discontinuous mode, defined with reference to the current in the transformer primary, which also doubles as an energy storage inductor.

The transformer winding polarity is such that no diode current flows when the transistor switches on. During this time energy is stored in the core. When the transistor switches off, the stored inductor energy causes the top of the secondary to develop a positive voltage which turns on the diode to charge the capacitor. During diode conduction, the transformer secondary appears as an inductor as shown in Chapter 10 for the boost circuit. Its stored energy keeps current flowing until the required amount of inductive energy is transferred to the capacitor. The control loop in the power supply adjusts the duty cycle of the power switch, such that the stored inductive energy matches that required by the load, thus keeping the average voltage on capacitor C_O constant.

Pertinent waveforms for each mode of operation are shown in Figure 186. The waveform set in part b is shown for the boundary condition; any further increase of transformer current will move operation into the continuous mode. Under idealized conditions, the peak diode current is given by:

$$I_{FM} = \frac{2I_0}{1 - D_{(MAX)}} - I_{O(MIN)} \qquad (11.5)$$

where: I_O = the dc output current

$D_{(MAX)}$ = the maximum transistor duty factor.

$I_{O(MIN)}$ = the minimum load current.

For a typical design in either mode, the boundary condition is the worst case. If $I_{O(MIN)} = 0$ and the maximum switch duty factor is 0.5 (a typical condition), $I_{FM} = 4\ I_O$. A low diode forward voltage is required for high efficiency.

As the transistor turns off, two steps of current commutation take place. First the collector current is commutated to the snubber capacitor C_S via snubber diode D_S. As the primary voltage continues to rise, it reaches a point where the rectifier diode D_R starts to conduct and current is commutated to D_R. During this second commutation interval, fast reverse recovery of D_S and low forward recovery of D_R are important to obtain efficient rectification. These effects not only cause losses, but they also produce a negative spike in the output voltage which increases ripple.

For the discontinuous mode, the decay of diode stored charge must be faster than the decay of current in the inductor. Selecting a diode whose $^*T_S$ is under 20% of the time from t_1 to t_2 usually yields satisfactory results. Should any charge be left in the diode when the transistor turns on, the circuit functions but the power switch experiences a reverse recovery current spike which lowers circuit efficiency.

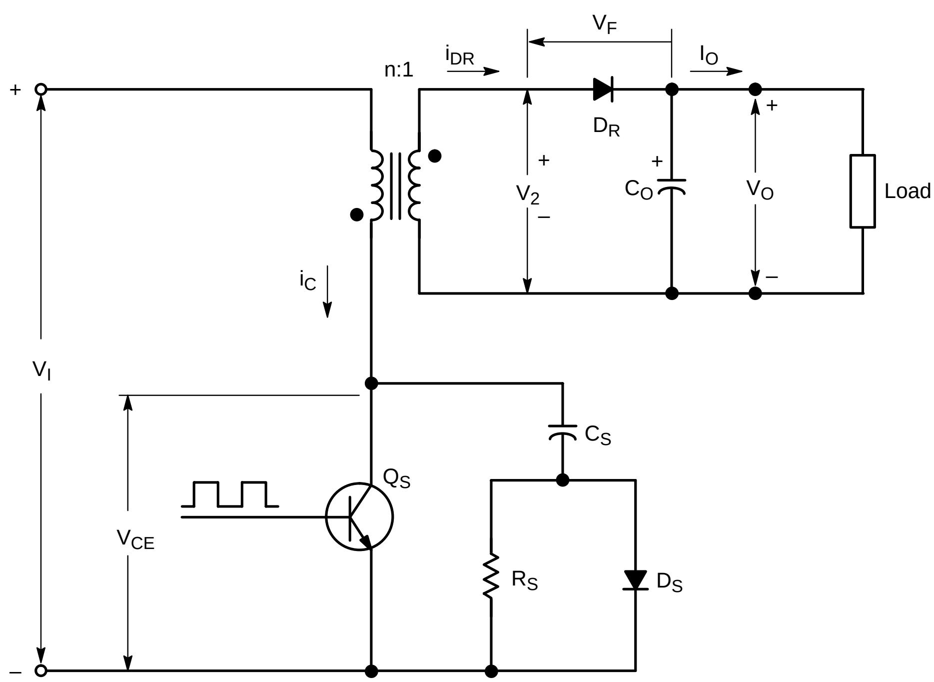


Figure 185. Output Section of a Flyback Converter Having a Transistor Snubber

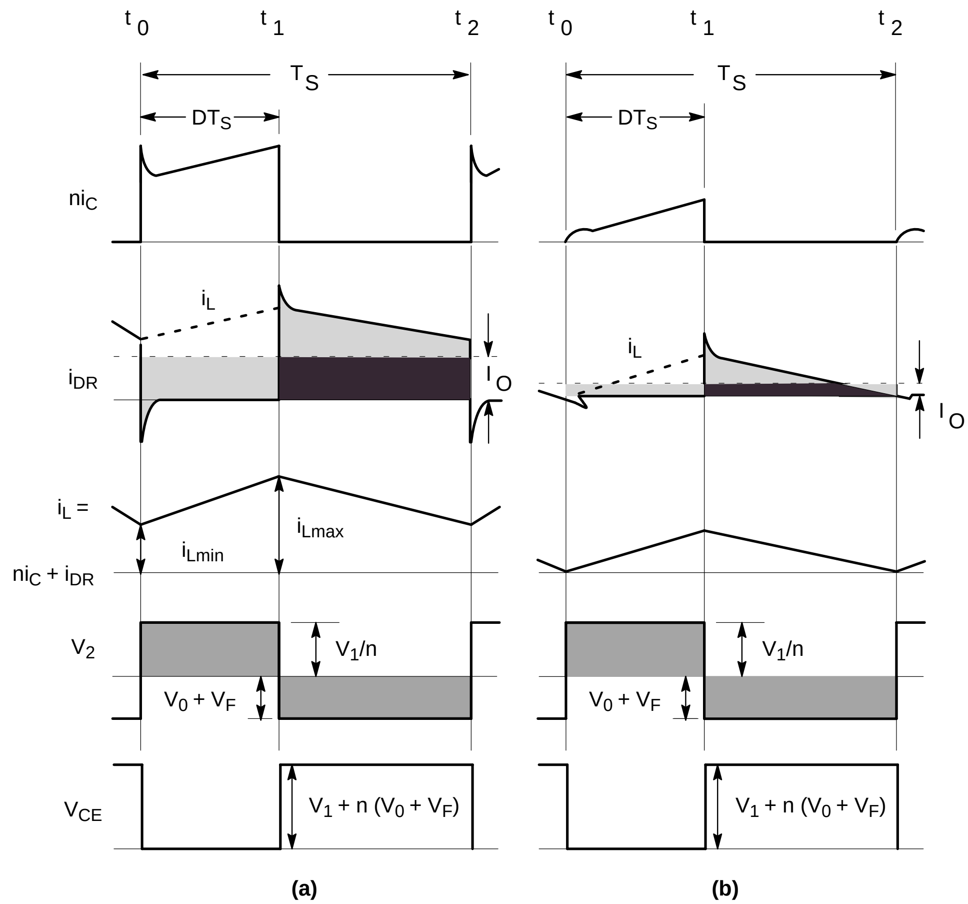


Figure 186. Flyback Converter Waveforms: (a) Continuous Mode; (b) Discontinuous Mode (Boundary Case)

In the continuous mode, significant diode current is commutated to the transistor when it turns on. The problems caused by hard switching – transistor turn–on loss and ringing – are reduced by using a fast diode. Nevertheless, a diode snubber is usually required (not shown on Figure 185).

The reverse voltage is the sum of the secondary voltage and the output voltage, plus any circuit ringing. The transformer is normally designed to produce the minimum required secondary voltage ($V_{FM} + V_O$) with a duty cycle of 0.5. At this point, the diode peak inverse voltage will only be a little over 2 V_O (transformer turns ratios never come out exactly as calculated and voltage drops must be considered). The worst case occurs when the converter operates at a high line condition. A 1.5:1 input voltage range is common; for this case, the minimum PIV will be three times the output voltage. Diodes selected for prototype work should have high voltage ratings, at least 1.5 times that predicted until the exact operating voltage conditions and parasitic transient voltages are determined in laboratory testing.

Forward Converter

The forward converter normally operates in the continuous mode and the diodes are subjected to forced commutation. A circuit diagram of the forward converter and associated waveforms are shown in Figures 187 and 188. Although a transistor snubber circuit is usually added, it is not shown because it does not influence diode behavior. The converter bears a superficial resemblance to the flyback converter just described but operation is entirely different. The transformer does not serve as a choke.

The winding phasing is such that when power switch Q_S is on, rectifier diode D_R is also on; current flow stores energy in the inductor L_O and feeds the load. When Q_S is turned off, the voltages on all windings reverse. In the secondary, D_R becomes reverse biased and the choke current is now carried by the freewheeling diode D_F. The center winding is used to reset the core by allowing the magnetizing current developed when Q_S was on to flow through the demagnetizing diode D_D to the supply voltage V_I. When Q_S is turned on, the winding polarities reverse to their former state and power is transferred to the load through D_R which commutates off the freewheeling diode D_F.

Transistor Q_S is turned on rapidly to reduce switching losses; consequently D_R turns on rapidly which causes a similarly rapid decrease of current in D_F and a potentially high reverse recovery current, which passes through D_R to the transistor. The peak current, whose rate of rise is limited only by transformer leakage inductance and transistor capability, could be significantly higher than the normal current at the end of the transistor conduction interval, if the diode is not suitably chosen. The current spike may cause a significant increase in turn–on dissipation in the transistor by raising its on–state voltage; the drive circuit could be designed to accommodate this current spike, but this solution lowers efficiency. In addition, the overcurrent protection scheme must be designed to ignore the spike. The lower the stored charge in the diode, the less difficulty is experienced with reducing turn–on losses and designing protection circuitry.

When the transistor turns off, several changes of state occur simultaneously. The end result is that the rectifier diode is cut off and the choke current flows through the freewheeling diode. When turn–off is initiated, the transistor current decreases rapidly compared to the current changes in the other components. Reset of the core through the demagnetizing winding and its series diode D_D commences. The demagnetizing current is a small fraction of the collector current. Also the rectifier diode current is commutated to the freewheeling diode. The reverse recovery current of D_R flows through the freewheeling diode causing a current spike during its forward recovery which increases its power dissipation. The recovery current of D_R also flows through the transformer and appears – modified by the turns ratio – as a reverse current through the demagnetizing diode D_D. The resulting reduction of current in D_D increases its impedance which in turn reduces the reverse recovery current. Thus the reverse recovery current is lower than might be expected, thereby prolonging the commutation interval in the secondary circuit.

Using ultrafast or Schottky diodes for the output diodes reduces the numerous problems associated with stored charge. The speed requirements for D_D are not severe; it operates in a soft recovery mode. It must, however, have a voltage rating over twice the maximum input voltage to the primary. The peak inverse voltage across the output diodes is equal to the secondary voltage minus the forward voltage across the conducting diode, which can be neglected when selecting diodes. The converter is usually designed such that the secondary voltage delivers that required to provide the output voltage when the input voltage is at its minimum level. At this point the peak inverse voltage is a little over twice the output voltage. The input voltage range is typically 1.5:1 and, allowing some margin for transient spikes, the voltage rating of the output diodes typically needs to be at least four times the output voltage. Since the output circuit switches by forced commutation, diode snubbers should be used.

Push–Pull DC–DC Converters

Popular push–pull configurations are shown in Figure 189. Although the stresses on the power switches differ, the voltage generated at the secondary of the transformer and the behavior of the rectifier diodes is identical for all three configurations. Push–pull converters can be viewed as two forward converters operating 180° out of phase and using a common transformer. Since the power switches alternately magnetize the core in opposite directions, no reset winding is required.

Because the secondary voltage waveform is alternating, a full–wave rectifier circuit is most efficient. Operation of the rectifier circuit is similar to the single–phase full–wave choke input circuit described in Chapter 7 where the inductor current is continuous. In a PWM power converter, a “dead time” occurs when both power switches are off causing secondary voltage to be zero. During this interval, each diode ideally carries half the choke current, while functioning as freewheel diodes.

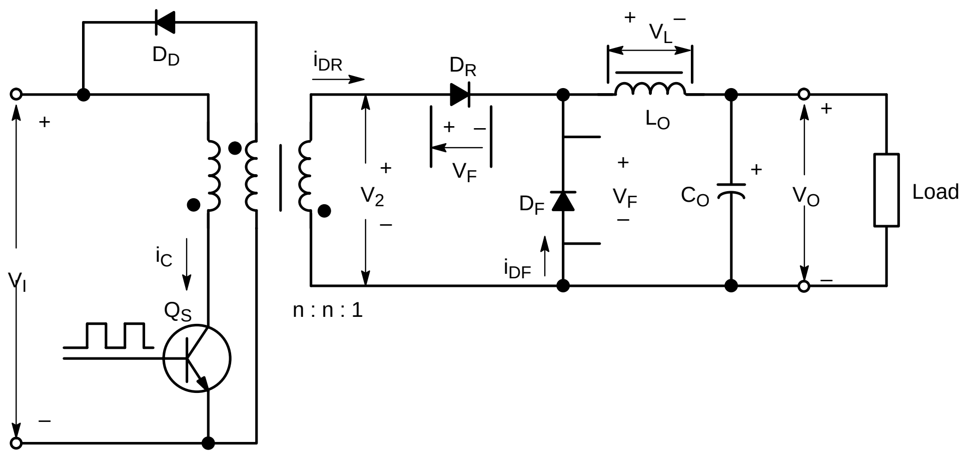


Figure 187. Output Section of a Forward Converter (Transistor Snubber not shown)

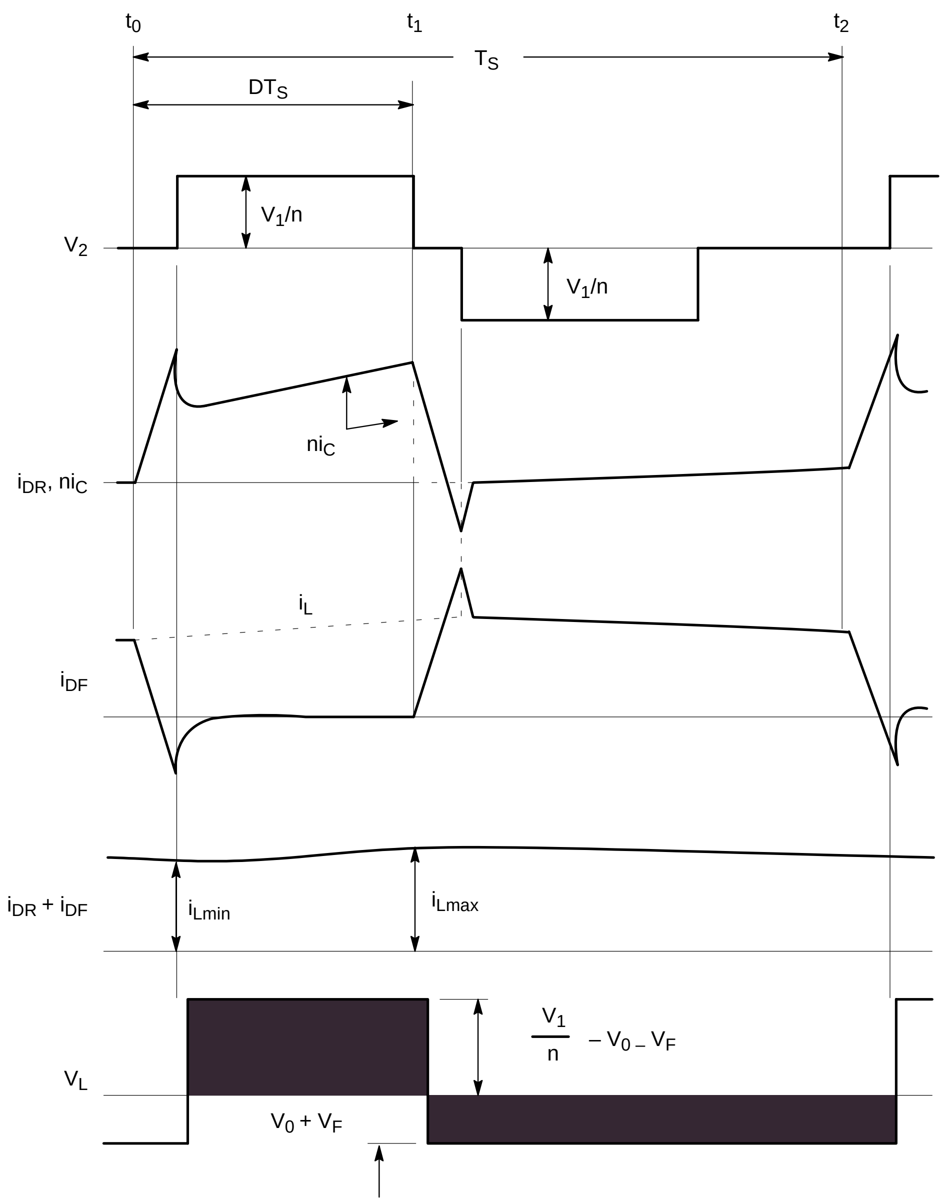
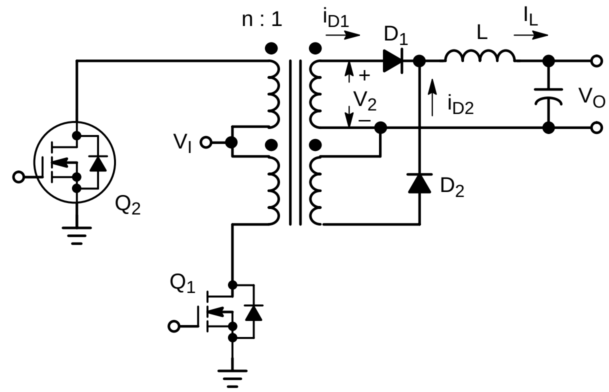
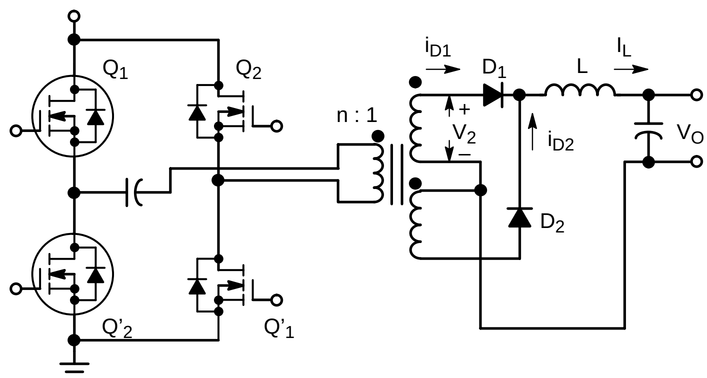


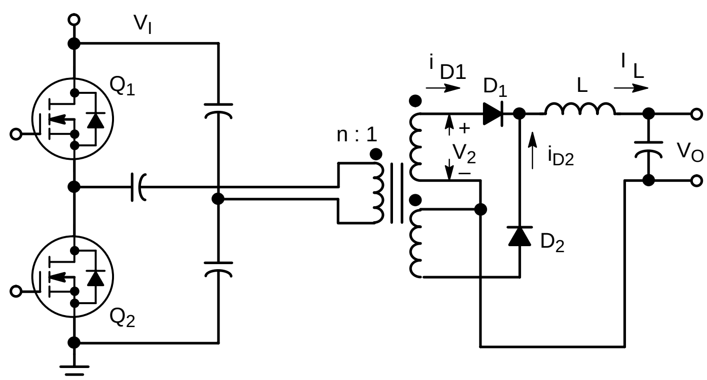
Figure 188. Forward Converter Waveforms for Continuous Mode Operation



(a)



(b)



(c)

Figure 189. Output Sections of Popular Push–Pull Converters: (a) Center–Tap; (b) Full–Bridge; (c) Half–Bridge

It is important that the diodes carry equal currents in order to develop equal but opposing flux in the transformer, thus avoiding core saturation. To aid in maintaining equal currents, the wiring runs to the diodes should be identical. Purchasing diodes in matched pairs is beneficial but often not worth the additional cost and complexity added to production.

The secondary voltage and resulting current in each diode are shown in Figure 190. The commutation intervals are indicated by the steep slopes on the diode current waveforms. Note that each diode experiences two distinct commutation intervals:

1. Current in one diode falls from full current to half current while the current rises from zero to half current in the other.
2. Current in one diode falls from half current to zero while current in the other rises from half current to full current.

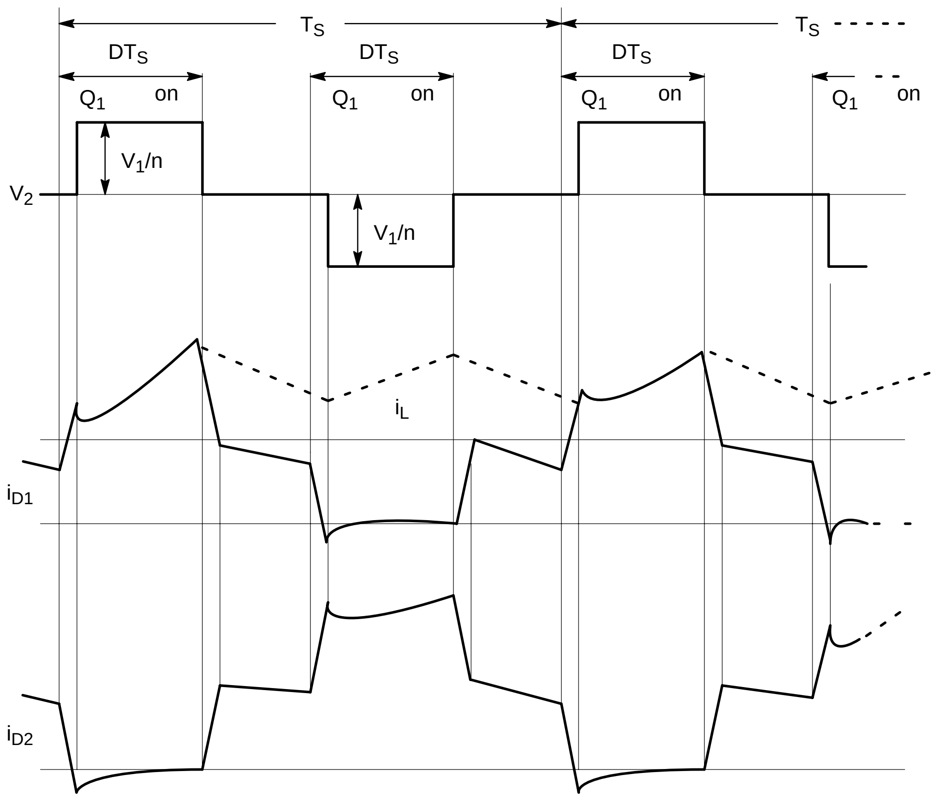


Figure 190. Secondary Waveforms for Push–Pull Converters

In interval 1, no reverse recovery spike occurs since the diode does not cut off. However, during interval 2, reverse recovery current occurs in the diode turning off; the reverse recovery current spike also appears in the current of the diode and the transistor, which are turning on. Because reverse recovery occurs when a diode is carrying only half the output current, the stored charge requirement is not as severe as for a forward converter operating at the same output level and diode switching frequency. Rectifier diode current handling requirements and voltage ratings are identical to the forward converter.

The MOSFETs in the circuits of Figure 189 show diodes connected between drain and source. These are part of the MOSFET, often called intrinsic or body–drain diodes. The area of the diode is large and its current handling capability generally exceeds that of the MOSFET, especially for high–voltage devices. Industry practice is to assign the diode a continuous current rating equal to the MOSFET's and a peak current rating three to five times greater.

Darlington transistors also have a built–in diode from collector to emitter. The Darlington's diode is the result of the emitter metallization partly overlaying the base to produce the monolithic base–emitter resistor of the output stage. Consequently, the contact and the collector base junction form a current path from the emitter to the collector. The total area of this structure and its current capacity need not be large to construct a properly functioning transistor, but they are purposefully enlarged to increase the collector–emitter diode's current handling capability. In general, low–voltage Darlingtons have diode current ratings equal to the collector current ratings, but in high–voltage Darlingtons, the diode current rating is typically half that of the collector current rating.

The utility of an intrinsic rectifier depends on the application in which it is used as well as the rectifier's characteristics. Circuits that use only one transistor rarely exercise internal diodes. Therefore, intrinsic diode characteristics are unimportant in "single–ended" buck, boost, flyback, and forward converters. On the other hand, in bridge circuits for power supplies or motor controls, the diodes are often used as circuit elements, where they function as catch or freewheeling diodes. As discussed, the catch function requires good forward recovery characteristics, while freewheeling requires good reverse recovery as well. Intrinsic diodes can fulfill necessary circuit requirements in many applications.

Off–Line AC–DC Converters

The proliferation of electronic equipment powered by switch–mode power supplies has created a great deal of difficulty for building electrical distribution systems [9]. The difficulties arise because of the narrow pulses of line current drawn by the capacitor input filters used to provide the dc power to the power switching stages. The current pulses have high rms values and a high harmonic content, which is fed into the electrical distribution system.

Among the problems caused are:

1. Circuit breaker life is short and improper operation occurs.
2. Excessive current flows in the neutral of the three phase power distribution system. The neutral is not supposed to carry current, so it is not protected by breakers; consequently, fires have occurred.

Recognizing the need to control harmonic currents, organizations such as IEC and IEEE are developing standards. When implemented, the stringent requirements of IEC standard 555–2 [10] will force manufacturers selling in Europe to develop new methods [11] of rectification and filtering for single–phase circuits rather than use the classic techniques described in Chapter 7.

One of the more popular schemes uses a continuous–mode boost converter operating at a high frequency (100 kHz, typically) under current mode control [11]. A simplified circuit is shown in Figure 191. The control circuit receives three inputs. A sample of the full–wave rectified sine–wave voltage is applied where it serves as a current waveform reference. A current transformer, as shown, or a resistor in series with the power switch provides a sample of the input line current. The third input is proportional to the dc output voltage. The controller uses these inputs to provide a PWM drive to the power switch. The drive achieves two objectives:

1. The dc output voltage is regulated by varying the ratio for switch on time to off time.
2. The average inductor current over one ac cycle (50 or 60 Hz) is forced to be sinusoidal by varying the on–time pulse width during the cycle such that the average of the pulses is sinusoidal.

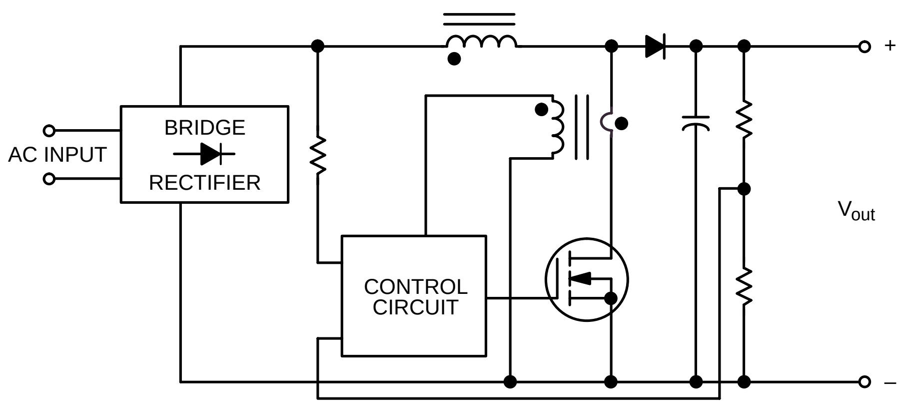


Figure 191. Popular High–Power Factor Rectification Circuit Using a Boost Converter

Rectifier Applications

The inductor and switch waveforms are sketched in Figure 192. Of course, in practice the repetition rate of the high–frequency switching pulses are much higher than shown and the pulses cannot be drawn to scale on a sketch of this kind.

The choke is usually operated in the continuous conduction mode over most of the ac cycle. At some point, typically around 30 V with a 120 V input, the inductor is allowed to "run dry" in order to keep it small. This creates a little step in the input current wave which raises its harmonic content. Thus a tradeoff between choke size and performance occurs.

The high–frequency rectifier diode D_R must meet stringent requirements. It operates in a forced commutated mode over most of the ac cycle so low stored charge is important. In addition, it must block the dc output voltage, which, for a boost converter, must exceed the highest line voltage peaks anticipated. For a "universal input" circuit, the boosted output voltage is typically around 400 V, dictating use of diodes with at least a 500 V rating. Fortunately, the current requirements are not severe since the current pulses have the same amplitude as would occur with a resistive load.

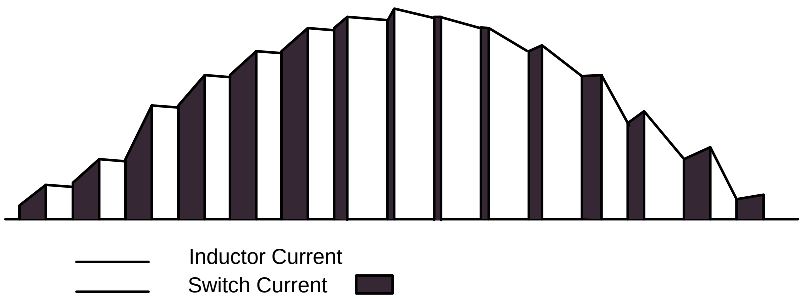


Figure 192. Inductor and Transistor Currents in the Boost Converter Rectifier

Synchronous Rectification

Three demands placed upon the power supply industry highlight the need for improved rectifiers:

1. Three–volt dc levels for the next generation of VSLI
2. Improved efficiency
3. Smaller size.

These demands are interrelated; without improvements in rectification efficiency, the efficiency of the 5 V supply is likely to remain about 80% and efficiency in a 3 V supply will be intolerable. Improvements in power density demand less internally generated heat.

A well–publicized method of improving rectification efficiency is to use a "synchronous rectifier." Although some innovative drive schemes have been developed which do not require additional windings on the power transformer [12], the most common embodiment consists of a low–voltage power MOSFET driven by an auxiliary winding on a power transformer as shown in Figure 193. The winding polarity is such that each MOSFET receives drive when it is to conduct current. Current flows from the MOSFET's source to its drain, which is opposite to that of conventional use. When the MOSFET turns off, it blocks drain–to–source voltage as it would in a standard application.

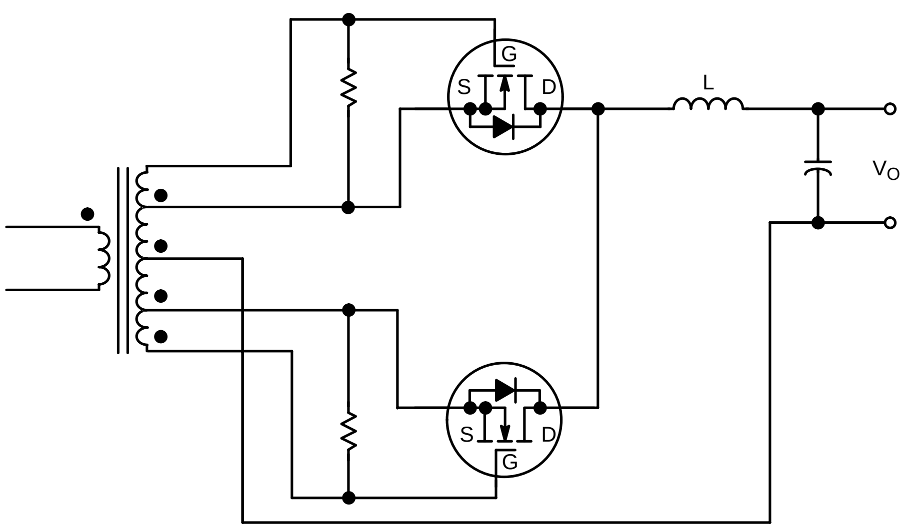


Figure 193. Rectification by Using Power MOSFETs as Synchronous Rectifiers

Synchronous rectification is possible because the MOSFET can conduct current in either direction when its channel is enhanced. Figure 194 illustrates the first and third quadrant operating characteristics of a 28 mΩ 60 V MOSFET. If the MOSFET's on–voltage is lower than the forward voltage of a comparable Schottky, rectification efficiency is improved. Designing a synchronous rectifier requires sequencing the MOSFET's turn–on. Some dead time is needed; otherwise, both transistors may be on simultaneously and shoot–through losses will be high. However, if the dead time is large relative to the carrier lifetime, the intrinsic diode will begin to conduct current, charge will be stored in its junction, and high reverse recovery losses will occur during commutation. A solution to the cross conduction problem is to use a Schottky in parallel with the MOSFET to conduct current during commutation, but the added cost and complexity makes this solution unappealing. Unique MOSFETs developed specifically for synchronous rectification have the best chance of challenging traditional rectification methods.

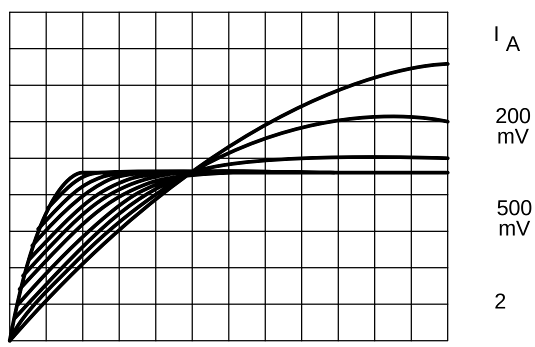


Figure 194. Power MOSFET Output Characteristics in the 1st and 3rd Quadrant

Synchronous rectification is not in widespread use at this time primarily because it is difficult to achieve lower loss than a Schottky provides. Simplicity of use is also strongly in favor of the Schottky. A major advantage of the Schottky is that its forward voltage decreases at elevated temperature, whereas the MOSFET's on–voltage increases. With present Schottky and MOSFET technologies, the Schottky is less expensive because its die size is smaller.

In spite of the Schottky's strengths, there are several applications where cost is of secondary interest that can benefit from the use of synchronous rectification. Sometimes the highest possible efficiency is required simply to save power, as in a battery–operated laptop computer where the operating ambient temperature is low [12]. In other situations, minimizing heat generation is essential because there is no outlet for it, as in some aerospace or military applications. Using a large MOSFET or a group of paralleled MOSFETs might be the only way to achieve the performance goals. Using synchronous rectifiers also provides new options for secondary side control and can help improve post regulation. The ability to control conduction time has enabled designers to omit ORing diodes in a system of paralleled converters [13].

References

1. Warren Schultz, “Baker Clamps–Traditional Concepts Updated for Third Generation Power Transistors,” *Powerconversion International*, July/August 1984. (Reprints available from ON Semiconductor Products Inc. Request AR 131.)
2. George Chryssis, *High Frequency Switching Power Supplies: Theory and Design*, McGraw–Hill Book Co., New York, N.Y. 1984.
3. Frank Cathell and Bill Roehr, “A Return from the Exotic–A Simple 1KW Off–Line Converter Using Two Power Switches and a Half–Bridge Topology,” *Official Proceedings of the Twelfth International SATECH Conference*, Boston, MA, October 25–31, 1986.
4. W. McMurray, “Selection of Snubbers and Clamps to Optimize the Design of Transistor Switching Converters,” *IEEE Transactions on Industry Applications*, Vol. IA–16, No.4, pp. 513–523, July/Aug 1980.
5. L.G. Meares, “Improved Non–Dissipative Snubber Design For Buck Regulators and Current–Fed Inverters,” *Proceedings of Powercon 9, Power Concepts*, Inc., 1982, B–2, pp.1–8.
6. Moshe Domb, “Nondissipative Snubber Alleviates Switching Power Dissipation, Second–Breakdown Stress, and V_{CE} Overshoot: Analysis, Design, Procedure and Experimental Verification,” *Conference Record of 1982 IEEE Power Electronic Specialists Conference*, pp. 445–454.
7. William J. Shaughnessey, “Modelling and Design of Non–dissipative LC Snubber Networks,” *Proceedings of Powercon 7*, March 24–27, 1980, San Diego, CA, pp. G4–1 to 04–9.
8. Sam Anderson, Kim Gauen, and William C. Roman, “Low Loss, Low Noise, MEGAHERTZ™ Rectifiers for High Frequency Applications,” *PCIM*, Oct. 1991.
9. “The Case for Power Factor Correction,” PCIM, December 1989, pp. 23, 24.
10. Hill Roehr, “A Synopsis of the Proposed IEC Harmonic Standard 555–2,” *PCIM*, June 1990, pp. 31.
11. Recent conferences have devoted entire sessions to the subject of power factor correction of line operated rectifiers. See: (a) *Conference Proceedings of 1990 and 1991 IEEE Applied Power Electronics Conference*, IEEE, New York, NY. (b) *The Proceeding of the Nineteenth International Power Conversion Conference*, 1989, Intertec Communications, Ventura, CA.
12. Bijan Mohandes, “Achieving 90% Efficient Power Conversion Using Synchronous Rectification,” *Proceedings of the Twentieth International Power Conversion Conference*, June 25–29, 1990, Munich, West Germany.
13. Ray On, “An Implementation of Synchronous Rectifiers in a Power Converter Designed for Use in Parallel Operation,” Proceedings of Power Conversion and Intelligent Motion ‘90, Philadelphia, PA, Intertec Communications, Ventura, CA.

General References

John 0. Kassakian, Martin F. Schlect, and George C. Verghese, *Principles of Power Electronics*, Addison–Wesley Publishing Company., New York, NY

Ned Mohan, Tore M. Undeland, William P. Robbins, Power Electronics: *Converters, Applications, and Design*, John Wiley and Sons, New York, NY.

Rudy Severans and Ed Bloom, *Modern DC–to–DC Switchmode Power Converter Circuits*,” Van Nostrand Reinhold Company, New York, N.Y., 1985.

Chapter 12

Reliability Considerations

Reliability Considerations

Reliability is a discipline that combines engineering design, manufacture, and test. An efficient reliability program emphasizes early investment in reliability engineering tasks to avoid subsequent cost and schedule delays. The reliability tasks focus on prevention, detection, and correction of design deficiencies, weak pans, and workmanship defects with the goal of influencing the product development process and producing products which operate successfully over the required life.

The reliability of a product can only be established if reliability is designed in, proper quality–controlled inspection and manufacturing assured; and maintenance procedures carried out. Improvement in reliability occurs when reliability information, obtained from test and field data, is analyzed and then, through an iterative feedback program, is used to improve the design.

This chapter discusses the fundamental reliability concepts and methods. The focus is on understanding the causes of failures in order to address them in design, manufacture, and test.

Concepts

Reliability is the probability that for a given confidence level, the product will perform as intended (i.e., without failure and within specified performance limits), for a specified mission time, when used in the manner and for the purposes intended under specified application environmental and operational conditions. Reliability is thus a quantitative metric used to assess performance over time, and can be considered to be a characteristic of a product in the sense that reliability can be estimated in design, controlled in manufacturing, measured during testing, and sustained in the field.

Quality describes the reduction of variability, so that conformance to customer requirements/expectations can be achieved in a cost–effective way. Quality improvement for a task or a process is described in terms of the target field, current status with respect to target (variability), reduction of variability (commitment to continuous improvement), customer requirements (who receives output, what are person's requirements/expectations), and economics (cost of nonconformance, loss function, etc.). While quality is a measure of variability (extending to potential nonconformances–rejects) for the product population, reliability is a measure of variability (extending to potential nonconformances–failures) for the product population, time, and environmental conditions.

Reliability serves as an aid to determine feasibility and risk when utilized early in the concept development stage of a product's development. In the design stage of product development, reliability analysis involves methods to enhance performance over time through the selection of materials, design of structures, choice of design tolerances, manufacturing processes and tolerances, assembly techniques, shipping and handling methods, and maintenance and maintainability guidelines. Engineering concepts such as strength, fatigue, fracture, creep, tolerances, corrosion, and aging play a role in these design analyses. The use of physics of failure concepts coupled with mechanistic as well as probabilistic techniques is often required to understand the potential problems and tradeoffs, and to take corrective actions when effective. The use of factors of safety and worst–case analysis are often part of the analysis. In the manufacturing and operations stage of a product, reliability analysis is useful in determining stress screening procedures, reliability growth, maintenance modifications, field testing procedures, and various logistics requirements such as the number of spare pans.

The reliability of electronic devices can be represented by an idealized graph called the bathtub curve. There are three important regions identified on this curve. In Region A, the failure rate decreases with time and it is generally called infant mortality or early life failure region. In Region B, the failure rate has reached a relatively constant level and is called the useful life region. Many modem semiconductors have been improved to the point where infant mortality and useful region failure rates are so near zero that the name bathtub curve is used as reference rather than expressing the true shape of the failure rate curve. Furthermore, the result of continuous reliability improvement is that many semiconductor products no longer require environmental stress screening such as bum–in.

In the third region, Region C, the failure rate increases again and is called the wearout region. Modem semiconductors should not reach the wearout portion of the curve when operated under normal use conditions (i.e., within the maximum rated value) and within reasonable lifetimes. In fact, the wearout portion of the curve, for many failure mechanisms, has been delayed beyond the useful life of much equipment. This is accomplished by physics of failure studies based on highly accelerated tests from which design rules are obtained which help to produce products with improved lifetimes.

Design for Reliability

Reliability in electronics design hinges on the recognition that an organized, disciplined, and time–phased approach to design, development, qualification, manufacture, and in–service management of a product is required to achieve mission performance over time, safety, support ability, and cost–effectiveness. The foundation of the approach consists of tasks, each having total engineering and management commitment and enforcement. The tasks are:

1. Define realistic system requirements.
2. Define the design usage environment (temperature, humidity, cycling ranges, vibration, etc., including all stress and loading conditions).
3. Characterize the materials, and the manufacturing and assembly processes (use design of experiments to determine and/or validate materials and processes).
4. Identify the potential failure sites and failure mechanisms (parts and part details, potential failure mechanisms and modes, architectural and stress interactions, measures for control).
5. Design to the usage and process capability (life cycle usage requirements, quality levels controllable in manufacturing and assembly).
6. Qualify the product manufacturing and assembly processes. (If all the processes are in control and the design is proper, then could move from product test, analysis, and screening, to process test, analysis, and screening.)
7. Control the manufacturing and assembly processes. (Processes must be monitored and controlled to avoid inappropriate shifts; this may involve screens and tests to assess statistical process control.)
8. Manage the life cycle usage of the product using closed–loop management procedures (realistic inspection and maintenance procedures, tracking procedures, deficiency reporting, and updates in procedures based on actual usage).

The reliability tasks which are utilized to design for reliability include system architecture and device specification, stress analysis, derating, stress control, stress screening, failure data collection and analysis, and failure modes, effects, and criticality analysis (FMECA). These are briefly discussed below.

System Architecture and Device Specification

As the physical design begins, reliability analysis can affect the system architecture and part selection, although functional and performance characteristics play the dominant role. Individual components must not be considered to be the only, or necessarily the major, source of failures. Interconnections and structures must also be properly selected. Furthermore, the interactions of stresses are important. Another aspect of system architecture is the use of redundant circuits. Redundancy may be deemed necessary for mission completion when the reliability estimates indicate improbable success or unacceptable risk.

Stress Analysis

Given the system architecture and pans, reliability prediction models are used to assess the influence of the magnitude and duration of the stresses on the reliability of the parts and systems, so that stress– and environment–controlling systems (i.e., vibration, cooling systems) and derating techniques can be implemented. Temperature, humidity, electrical fields, vibration, and radiation are major stress variables affecting reliability.

There are various ways in which both the operating and environmental stresses can be controlled to improve reliability. Methods can be applied to keep harmful stresses (i.e., high temperatures, high shock loads, high humidity, high radiation etc.) away from sensitive devices and structures, or to manage the system environment to obtain controlled stress conditions. Lowering the harmful stresses is often a first choice of designers for reliability improvement. However, the cost and complexity of lowered stresses must be balanced against the cost and complexity of electronic complications to improve reliability by improved architectures and pans.

Derating

Derating is based on the concept that operating electrical, thermal–mechanical, and chemical stresses accelerate failures in a predictable manner which, if controlled, will improve reliability. Using the mathematical expressions of reliability prediction, one can often derive a derate schedule. Such schedules must be based on the dominant failure mechanisms for the particular electronics and must include interconnects and device interactions, as well as the devices themselves.

Stress Screening

Screening is the process by which defective parts, resulting from improper or out–of–control manufacture and assembly processes are detected and eliminated from a production batch. The principle involves observing or inducing latent defect failures only in a population of parts that already has “weak” parts without reducing the reliability in the population of “strong” parts. The assumption is that, through the application of short–term stresses, the weak population can be discovered and eliminated, leaving a highly reliable population. Stress screening methods and the associated acceleration stress levels must be based on the failure mechanisms due to problems in quality.

One type of screen is called bum–in, whereby the parts are operated for a period of time at high temperatures in order to precipitate defects in the weak population of pans. For parts with low failure rates (below ten failures per million device hours) ON Semiconductor believes bum–in prior to usage does not remove many failures. On the contrary, it may cause failures due to handling.

Failure Modes, Effects, and Criticality Analysis (FMECA)

FMECA is a method to assess the interoperability of the pans, subassemblies, assemblies, and subsystems comprising the system. The objective is to determine the effect of failures on system operation, identify the failures critical to operational success and personnel safety, and rank each potential failure according to the effects on other portions of the system, the probability of the failure occurring, and the impact of the failure mode. Reliability predictions are often used for each element in the system for each potential failure mode.

Failure Data Collection and Analysis

Failure data comes from life tests, accelerated life tests, any environmental stress screens that may have been performed, and, if available, field characterization and malfunction data. Users of this data must be aware of the source of data, the specific failure modes uncovered in failure analysis, the manufacturing and handling conditions, the application of the parts in the field, and the applied stress history. Test data often neglect the kinds of stresses that actually arise in the field, while field data collection is notoriously inadequate, in the sense that a failure is often associated with a removal rather than with a well–identified failure mechanism. In other cases, the cause of failure is either not reported or inaccurately reported if it was caused from being installed incorrectly, mishandled, or misused. Users of data must be especially cautious when using models extrapolated from field– and test–based models for new technology–type pans having very little, if any, application data. Failures should be classified by the failure location or failure site, the consequence or mode by which the failure is observed (i.e., open, short, parameter drift) and the failure mechanism (i.e., electronic overstress, corrosion, electromigration, fatigue).

Practical Aspects of Semiconductor Reliability

An understanding of semiconductor reliability requires knowledge of the potential failure mechanisms and the stresses that accelerate failures. What follows is a brief overview of each of the various failure mechanisms affecting semiconductor devices. A focus is placed on temperature dependency because traditionally steady–state temperature has incorrectly been considered to be a critical parameter to reliability. This section shows that below rated temperatures, steady–state temperature is not a major contributor to failures and that more attention should be paid to temperature cycling and temperature gradients. Not all mechanisms described are present in all power rectifier diodes, but for complete coverage of the topic, are included here.

Electrical Overstress

Electrical overstress occurs when a higher–than–rated voltage or current induces a hot–spot temperature beyond specifications for short periods of time. Hot–spot development typically occurs at a semiconductor junction due to an increase in the current flow to accommodate the additional stress in the device. As the junction heats up, the increased temperature encourages increased current flow due to the lowering of the silicon resistance at higher temperature which in turn further heats up the junction. If the situation continues, the hot–spot temperature may exceed the intrinsic temperature of silicon, beyond which the resistivity of silicon decreases greatly. This allows more current to pass through the hot spot, further increasing the temperature, and resulting in thermal runaway. Failure occurs when the silicon in the hot spot melts, destroying the silicon crystal structure. If the electrical transient continues, the interconnects are expected to melt as well.

Second Breakdown

When reduced to basics, excessive temperature causes most failures in semiconductor chips. High uniform temperature within the chip is responsible for a gradual drift of characteristics that becomes more noticeable as temperature is raised, but this rarely causes catastrophic failure, such as opens or shorts.

Catastrophic failures result from extreme temperature in a lead wire causing it to open, or from a hot spot in the chip causing it to short. In most circuits, the latter also causes the former. A reverse voltage is always required in order to produce a sufficient amount of current crowding to cause a hot spot, and high hot–spot temperatures cause an extreme reduction in the voltage capability of a semiconductor. A thermally induced breakdown, observable on a V–I plot, occurs as shown in Figure 195.

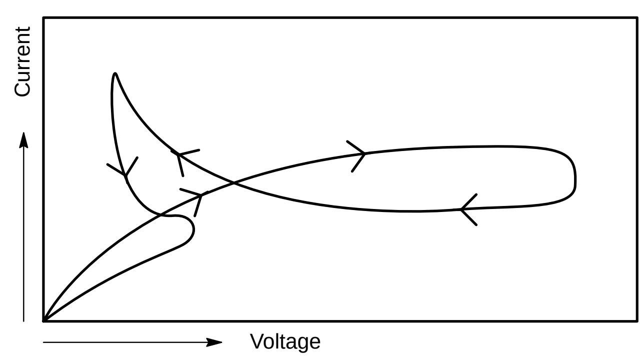


Figure 195. Typical V–I Plot of Semiconductor Material Exhibiting Thermal Breakdown

Thermal breakdown – a switch back of voltage from a high to low level – has been shown to be a fundamental property of any semiconductor material. It occurs at the temperature at which the material becomes intrinsic. The temperature of intrinsic conduction is inversely related to resistivity, varying from 200°C to 400°C for resistivity values in common use.

The most thermally rugged devices are silicon, diffused–junction, or Schottky barrier semiconductors of low resistivity. Silicon is better than germanium because it becomes intrinsic at higher temperatures; a diffused junction results in more even current distribution than does an alloyed junction, and the lower the resistivity of the material, the higher the temperature at which it becomes intrinsic.

Ionic Contamination

Ionic contamination causes reversible degradation such as a threshold voltage shift and gain reduction in MOS devices due to the presence of mobile charge ions within the oxide or at the device–oxide interface. Ionic contamination is caused by mobile ions in semiconductor devices. Contamination can arise during packaging and interconnect processing, assembly, testing, screening, and during operation. Alkali ions, such as sodium, chloride, and potassium, are the most common contaminants. The amount and distribution of the alkali ions in or near the gate dielectric region will influence the device threshold voltage by superposition of these ionic charges on externally applied device voltage. Extra positive charge at the Si/SiO_2 interface induces extra negative voltage resulting in a decrease in the threshold voltage of the device. The mobility of the ions is temperature dependent. For this reason, a high–temperature storage bake and exposure to high temperature during bum–in are found to screen out the ionic contamination failures. A common characteristic of contamination failures is the reversibility under high temperatures in the neighborhood of 150 to 200°C which partially or fully restores device characteristics, reflecting a disordering of charge accumulations resulting from ion mobility and applied bias.

The problem of ionic contamination is solved by use of high–purity materials and chemicals for processing, use of HCl during oxidation, and use of phosphosilicate glass (PSG) or borophosphosilicate glass (BPSG) over the polysilicon to getter the ions. Passivation layers can provide additional protection against ingress of alkali ion contamination in completed devices.

Electromigration

Electromigration is the result of high current density, typically of the order of 10^6 A/cm^2, in metallization tracks which produces a continuous impact on the metal grains causing the metal to pile up in the direction of the current flow and produce voids in an upstream direction with respect to the current flow. In thin–film conductors, electromigration–induced damage usually appears in the form of voids and hillocks. Voids can grow and link together to cause electrical discontinuity in conductor lines which leads to open circuit failure. Hillocks can also grow and extrude out materials to cause short–circuit failure between adjacent conductor lines in the same level or in adjacent levels in multilevel interconnecting structure. Alternatively, it can break through the passivation or the protective coating layers and lead to subsequent corrosion–induced failures. Void–induced open–circuit failures usually occur at an earlier time than extrusion–induced short–circuit failure in thin–film conductors using Al–based metallurgies.

The phenomenon of electromigration–induced mass transport is attributed to atomic flux during the passage of current through a polycrystalline thin–film conductor, as a result of electromigration in the lattice and at the grain boundaries. There is a temperature gradient dependence of electromigration failures. An analysis of electromigration tests on the lifetime of Cr/Al–Cu conductors covered with polymide passivation showed that the electromigration failure location is typically nearer to the location of maximum temperature gradient, whereas the location of a nontemperature–gradient–induced failure was randomly distributed.

Hillock Formation

Hillocks in die metallization can form as a result of electromigration or exposure to extended periods under temperature cycling conditions. Hillock formation as a result of electromigration often occurs upstream in the sense of the electron flow, from the area of voiding. Hillocks grow fastest at the down–stream edge, closest to the source of migrating metallization. Both voiding and hillock formation can occur, sometimes on top of each other. This phenomenon of simultaneous voiding and hillock formation occurs at temperatures in the neighborhood of 140 to 200°C.

Hillock formation produced by extended periods under temperature cycling conditions is believed to be due to a self–diffusion process that occurs in the presence of strains within the metallization. These strains may be due to a mismatch in the thermal expansion coefficients of the metallizations (Au), underlying refractory layer (TiW), Si, and SiO_2. Hillock growth is more extensive in films deposited on room–temperature substrates than on heated substrates, indicating that the effect may vary with grain size. Hillocks can cause electrical shorts between the adjacent lines and fracture of the overlying dielectric film. In double–level metallized devices, hillocks can result in shorts between the underlying metal layer and the overlying metal layer. Hillocks can be caused by electromigration. They can result in thin dielectric sites that are susceptible to subsequent breakdown of the intermetal dielectric. Hillock formation is largely remedied by use of alloyed metal such as Al–Cu and improvements in the technologies of passivation and packaging. Hillock formation is more a function of temperature cycling and temperature gradients and is mildly dependent on temperature in the neighborhood of 400°C.

Contact Spiking

Contact spiking is the penetration of metal into the semiconductor in contact regions at temperatures typically above 400°C. This failure mechanism is accompanied by solid–state dissolution of the semiconductor material into the metal or the alloy of metallization system with the metal semiconductor interface moving vertically or laterally into the semiconductor. Contact spiking in chips is observed at high chip temperatures or localized high contact temperatures. Localized high temperatures may cause failure of the chip to substrate bond, thus increasing the thermal resistance or producing a thermal runaway of the device, or resulting in a large degree of electrical overstress.

Contact spiking can occur in device fabrication when the device is exposed to high temperatures. This failure mechanism can be minimized by using silicon containing aluminum alloys such as Al–1% Si or using barrier metals such as Ti–W. Metal spikes penetrate into the electrically active region of the device causing increased leakage current or shorting. It is not possible to characterize such interdiffusion failure mechanisms by an activation energy because of their irregular behavior.

Migration of Al along Si defects has been observed in NMOS LSI logic devices, due to contact migration also known as electrothermomigration. Failures are accelerated by elevated ambient temperature in the neighborhood of 400°C. This failure mechanism may be dominant during VLSI manufacture and packaging when the temperatures exceed 400°C.

Metallization Migration

Metal migration occurs between biased lands under conditions conducive to electro crystallization. Dendritic growth is a common cause of failure. Conditions for metal migration include:

1. A level of current density at the tip of the dendrite in the neighborhood of 10 A/cm^2.
2. Spheroidal and parabolic diffusion.
3. Sufficient liquid medium such as condensed water.
4. Applied voltage which exceeds the sum of anodic and cathodic potentials in equilibrium with the electrolyte
5. Materials with defects which allow water condensation to satisfy the current density requirement.

Metallization migration involves the formation of aluminum growths beneath the silicon dioxide layer. Metallization migration occurs during deposition of conductors on silicon dice due to the combined effects of elevated temperature and electrical stress. It has been reported that triangular–shaped aluminum growths formed beneath the silicon dioxide layer when conductor metallization was deposited at high temperature on silicon substrate. The triangles formed within minutes after deposition of an 8000 angstrom thick aluminum layer on silicon at temperatures from 500 to 577°C. The time for growth formation decreases with increasing temperature. Growth caused local electrical short circuits. The high temperatures at which this failure phenomenon occurs makes it a recessive mechanism in the normal operation of microelectronic devices.

Corrosion of Metallization and Bond Pads

Corrosion is typically defined as the chemical or electrochemical reaction of a metal with the surrounding environment. The mechanisms of corrosion can be divided into two types: dry, such as oxidation of aluminum in air and wet, where the reaction occurs as in the presence of an electrolyte, a moist environment, and an electromotive force. Dry corrosion is of minor importance in semiconductor devices, since the corrosion process is self passivating for the metal, fanning a thin oxide film which prevents further oxidation. Wet corrosion, in the presence of an ionic contaminant and moisture, can provide a conductive path for electrical leakage between adjacent conductors, dendrite growth, or corrosion of device metallization or bond pads. Corrosion typically begins when the temperature inside the package is below the dew point, enabling condensation inside the package. During operation, heat dissipated by the die is often enough to raise the temperature to evaporate the electrolyte. Elevated temperature due to device power thus acts as a mechanism to inhibit the corrosion process.

Stress Corrosion in Packages

Stress corrosion is a failure phenomenon which occurs around 300°C and above, predominantly in power devices. Electrochemical corrosion interacting with mechanical stress is a potential cause of failure. Failures are typically transgranular and result from the acceleration of the fatigue process by corrosion of the advancing fatigue crack. For example, craze cracking of polymers occurs in halogenated solvent vapors. Adsorbed moisture films due to high relative humidity exposure cause static fatigue of the lead zinc–borate sealing glass. The distinct mechanism is chemical attack on glass, and the time to failure is usually modeled by an Arrhenius equation of temperature which activates above 300°C. The rate of galvanic corrosion in the presence of a liquid electrolyte increases as temperature is increased due to a more rapid rate of electron transfer. Typically corrosion products in microelectronic devices, such as $Al(OH)_3$, are derived from reaction processes which are monotonically increasing functions of temperature. Although the corrosion rate depends in part on the steady–state temperature, corrosion rate also depends on the magnitude and polarity of the corrosion galvanic potential which are functions of the electrolyte concentration, pH, local flow conditions, and aeration effects. Finally, as noted previously, elevated temperature due to component operation and especially continuous high power, acts as means to inhibit corrosion.

Wire Fatigue

The wires used to connect die bond pads to leads can fail due to cyclic temperature changes resulting from repeated flexing of the wire. The most prevalent failure site is the heel of the wire. Wire–bond failures are typically due to the differential coefficient of thermal expansion between the wire and the package as the device is heated and cooled down during temperature and power cycling. Failures may be inhibited by high loops, though small dimensions of present–day packages impose stringent requirements on the loop dimensions, making this solution often impractical.

Wire Bond Fatigue

A wire bond subjected to temperature change experiences shear stresses between the wire, the bond pad, and the substrate as a result of differential thermal expansion. Intermetallic brittleness and growth is enhanced by temperature cycling. When bonded together, gold and aluminum present a problem when subjected to temperature cycling. Au–Al intermetallics are more susceptible to flexure damage than pure Au and Al wires. Kirkendall voids occur at the interface between the wire and the bond pad when the contacting materials are gold and aluminum. Kirkendall voids form when either the aluminum or gold diffuses out of one region faster than the other can diffuse from the other side of that region. Vacancies pile up and condense to form voids, normally on the gold–rich side along the gold–to–aluminum interface.

The rates of diffusion vary with temperature and are dependent upon the adjacent phases, as well as on the number of vacancies in the original metals. The compounds formed in the intermetallic are often called purple plague because of the purple appearance of the gold–aluminum compound. Studies show that the formation of significant amounts of purple plague is negligible below 150°C. Intermetallic compounds are mechanically strong, brittle, and electrically conductive. Temperature and power cycling can cause failure as a result of differential thermal expansion between the intermetallic and the surrounding metal, and reduce mechanical bond strength as a result of voiding of the surrounding metal, usually accompanying intermetallic formation. Gold–aluminum intermetallics are stronger than pure metals and their growth is enhanced by increased temperature and temperature cycling.

Generally, the bond strength is more a function of temperature cycling than steady–state temperature in the range of temperature between –55°C and 125°C, although the bond strength decreases as a function of temperature above 150°C for gold–aluminum bonds and above 300°C for gold–copper bonds.

Die Fracture

The die, the substrate or the leadframe, and the case of a package typically have different thermal expansion coefficients. For example, die are usually made of silicon while the substrate is typically alumina, berylia, or copper having a coefficient of thermal expansion different from the die material. As the temperature cycle magnitude rises during temperature and power cycling, tensile stresses are developed in the central portion of the die and shear stresses are developed at the edges of the die. Ultimate fracture of the brittle die can occur suddenly without any plastic deformation when surface cracks at the center of the die or at the edge of the die reach their critical size and propagate during thermal cycling to the critical crack size.

Vertical cracking of the die is caused by tensile stresses and horizontal cracking is caused by high shear stresses at the edges. The brittle failure criterion can be represented by the size of a critical crack on the external die surface. Microcracks are developed in the die during manufacture. In some cases these microcracks may be large enough to cause brittle failure of the die. A pre–existing defect may develop into a crack under the influence of thermal cycling in the die. This crack may not be of critical size at the applied service stress, but may grow to critical size gradually by stable fatigue propagation. The rate of fatigue crack propagation per cycle is determined by the cyclic change in stress intensity factor, a measure of the stress at or around the crack tip. The stress generated in the die during temperature cycling is dependent on the number and magnitude of the temperature cycles.

Die Adhesion Fatigue

In a typical power or environmental temperature cycle, the die, the die attach, the die substrate, and the package experience temperature differences and temperature gradient differences. Because they also have different coefficients of thermal expansion, the bond between the die and the substrate can experience fatigue failure.

The most common die attach defects are voids. The presence of edge voids in the die attach induces high longitudinal stresses during power and environmental temperature cycling. These voids can act as microcracks which propagate during power cycling resulting in debonding of the die from the substrate or the substrate from the case. Voids are responsible for weak adhesion, die lifting, increased thermal resistance, and poor power cycling performance. Solder and some epoxies have been shown to crack and detach during temperature cycling. Voids can form from melting anomalies associated with oxides or organic films on the bonding surfaces, out gassing of the die attach, trapped air in the bonds, and shrinkage of solder during solidification. Insufficient plating, improper storage, lack of cleaning, or even diffusion of oxidation prone elements from an underlying layer can generate voids during melting of die attaches. In other instances, dewetting of solder results in excessive voiding, especially when a solderable surface, a poor solderable underlying metal, or excess soldering time produces an intermetallic compound not readily wet by the solder. Even under ideal production conditions, voids are often present due to solvent evaporation or normal outgassing during cooling of organic adhesives. Although voids can form from a number of sources, they are normally limited to an acceptable level

through process control. The actual package construction, the die attach unit materials, and the overall void concentration determine the actual effect of voiding on device reliability. While small concentrations of random voids have little effect on the peak junction temperature, a large–size void may reduce the thermal performance of the device by creating a large temperature gradient.

Cracking in Plastic Packages

Thermal coefficient of expansion (TCE) mismatches drive the failure mechanisms of cracking of plastic packages. Silicon has a TCE of about 3 ppm/°C, while molding compounds typically have a TCE of about 20 ppm/°C below the glass transition temperature. Most lead frame materials are either Alloy 42 with a TCE of 4.7 ppm/°C or copper with a TCE of about 17 ppm/°C.

A typical process involves die attachment with adhesive cure at 270°C using polyimide adhesive or at l70°C using epoxy adhesive, followed by encapsulation and cure of the molding compound at 170°C. Negligible stress is established at these elevated temperatures, but the thermal coefficient of expansion differentials causes increasing stress as the ambient temperature is lowered. The die attach produces a bending moment at room temperature that puts the surface of the die into tension and the bottom of the die into compression. The magnitude of the tensile stress at the surface of the die is proportional to the thickness of the die. Encapsulation superimposes a compressive stress on the die, and tends to place the molding compound under tension. Cracks therefore have a tendency to propagate in the molding compound.

Cracks in packages are a reliability concern because they provide a path for the entry of contaminants. Shear stresses are applied on the surface of the chip due to differential thermal expansion between the plastic encapsulant and the passivation layer, resulting in the fracture of the passivation layer, which forfeits its action as an impurity getter and barrier. This also causes lateral displacement of the interconnection lines on the surface of the chip and shorts between the overlying and underlying interconnection patterns in devices with two or more layers of interconnection. The effects of stresses applied by plastic are more pronounced in larger chips especially at the chip corners. Resilient conformal coating on the die and the use of plastic encapsulants with a lower coefficient of thermal expansion are some of the methods used to decrease thermal stress effects in the passivation layer, due to thermal coefficient mismatches. Damage of passivating film of plastic encapsulated devices can occur due to thermally induced stress at the encapsulant passivation layer interface. The location of aluminum corrosion coincides with defects in the passivating film, typically caused by differential thermal stress from the encapsulant. Cracks also initiate due to stress concentrations at the top edge of the die or at the bottom edge of the lead frame, or due to voids in the molding compound. The cracks propagate under the influence of temperature cycling.

Delamination is a failure mechanism that occurs between the die and molding compound and between the molding compound and the lead frame due to shrinkage after molding. During board assembly, plastic surface mount components (PSMCs) are subjected to temperatures in the range of 215–260°C and heating rates of as high as 25°C/sec during normal operational temperature cycles. In addition to TCE driven stresses, if the heating or cooling rates are much higher than 10°C/sec, thermal stresses associated with internal temperature gradients can be appreciable. Moisture absorbed by the package molding compound evaporates under high temperatures, creating an internal pressure causing cracking (generally referred to as the popcorn mechanism). It is possible for PSMCs to absorb moisture during storage in normal ambients for vapor–driven stresses to dominate the mechanics of the package. The main factors influencing the propensity to crack include: peak temperature reached during soldering, moisture content of the molding compound, dimensions of the die, thickness of the molding compound under the die, and the adhesion of the die to the lead frame. Baking prior to soldering or the use of porous encapsulants eliminates this problem.

Design for Reliability

In general, about the only reliability–improvement measure available to electronic equipment level designers is to derate. As mentioned earlier, derating is a technique by which either the operational stresses acting on a device or structure are reduced relative to rated strength, or the strength is increased relative to allocated operating stress levels. For example, manufacturers of rectifiers specify limits for blocking voltage, forward current, power dissipation, and junction temperature. The equipment designer has the choice to select an alternative component, or make a change in the operating conditions (i.e., derate a specific parameter, such as temperature, below the rated level).

The derating factor, typically defined as the ratio of rated level of a given “stress” parameter to its actual operating level, is actually a margin of safety or a “margin of ignorance.” The amount of derating is determined by the impact that a failure could cause, and by the amount of uncertainty inherent in the reliability derating model and all its inputs. Ideally, this margin should be kept to the minimum required, so as to maintain the cost–effectiveness of the design. This puts the responsibility on the reliability engineer to identify, as unambiguously as possible, the rated strength, the relevant operating stresses, and the reliabilities.

Steady–state temperature has typically been assumed to be the easiest parameter to derate. However, only a few of the failure mechanisms are strongly dependent on steady–state temperature in the range of operating temperatures between –55°C to 150°C, and the use of the Arrhenius temperature equation as a generic relationship is misleading.

Examinations of case histories indicate that temperature's relationship to reliability may be more a case of exposing incompatibilities between operational requirements and design or manufacturing processes. In other words, the product as designed is not suitable for operation in the desired environment without changes. In such cases certain questions must be addressed before taking action:

- Will lowered temperature by itself avoid the experienced failures? If so, how much should it be lowered and how is the conclusion reached?
- Will higher temperature by itself accelerate the experienced failures? If so, how much should it be raised and how is the conclusion reached?
- Will lowering the magnitude of temperature change avoid the experienced failures? If so, how much should it be lowered and how is the conclusion reached?

Each action has its own costs, and the designer or user must fully understand the implications of each to produce cost–effective, reliable equipment. Often, temperature is judged to be the culprit, and blaming temperature may shift emphasis away from actual design or manufacturing inadequacies. In some cases where temperature may be identified as an important ingredient in the explanation for unreliability, the product as designed is not suitable for operation in the desired environment without being changed.

To be effective, derating criteria must target the right stress variables to address a given failure mechanism, clearly specify the inputs to failure models, and recommend tests to measure the parameters accurately. Physics of failure concepts can serve to relate allowable operating stresses to design strengths, through quantitative modeling of the relevant failure mechanisms. Field measurements may also be used, in conjunction with modeling simulations, to identify the actual operating stresses at the failure site. Once the failure models have been quantified, the impact of derating on the effective reliability of the component for a given load can be determined. Quantitative correlations between derating and reliability enable designers and users to tailor the margin of safety more effectively to the level of criticality of the component, leading to better and more cost–effective utilization of the functional capacity of the component.

Reliability Testing

The following summary gives a brief description of the various reliability tests. These tests are commonly performed by manufacturers to ensure that a product is properly designed and assembled. Quality products do not fail in the field from these environmental conditions. Not all of the tests listed are performed by each product group, and other tests not described are sometimes performed when appropriate. In addition, some form of preconditioning may be used in conjunction with the following tests.

High–Temperature Operating Life

High–temperature operating life testing is performed to accelerate failure mechanisms that are thermally activated through the application of extreme temperatures at biased operating conditions. The temperature and voltage conditions used will vary with the product being stressed. All devices used in the test are sampled directly after final electrical test with no prior burn–in or other prescreening unless called out in the normal production flow. Testing can either be performed with dynamic signals applied to the device or in static bias configuration. Common failure modes include parametric shifts and catastrophic events. Common failure mechanisms include foreign material, crack and bulk die defects, metallization, and wire and die bond defects.

Temperature Cycle

Temperature cycle testing accelerates the effects of thermal expansion mismatch among the different components within a specific die and packaging system. During temperature cycle testing, devices are inserted into a cycling system and held at the cold dwell temperature long enough to incorporate thermal stabilization and appropriate creep and stress relaxation of the materials. Following this cold dwell, the devices are heated to the hot dwell where they remain for another minimum time period. The system employs a circulating air environment to assure rapid stabilization at the specified temperature. The dwell at each extreme, plus the two transition times, constitute one cycle. Test duration for this test varies with device and packaging system employed. Typical test duration is for 1000 cycles, with some tests extended to look for longer–term effects. Common failure modes include parametric shifts and catastrophic events. Common failure mechanisms include wire bond, cracked or lifted die, and package failure.

Thermal Shock

The objective of thermal shock testing is similar to that for temperature cycle testing. However, thermal shock provides additional stress in that the device is exposed to a sudden change in temperature due to the rapid transfer time. Failure mechanisms caused by temperature transients and temperature gradients can be detected with this test. Devices are typically placed in a fluorocarbon bath and cooled to a minimum specified temperature. After being held in the cold chamber for some minimum period of time, the devices are transferred to an adjacent chamber filled with fluorocarbon at the maximum specified temperature. Test duration is normally for 1000 cycles with some tests being extended to look for longer–term effects. Common failure modes include parametric shifts and catastrophic events. Common failure mechanisms include wire bond, cracked or lifted die, and package failure.

Temperature/Humidity Bias (THB)

Temperature/humidity bias (THB) is an environmental test designed to measure the corrosion/moisture resistance of plastic encapsulated circuits. A nominal reverse bias is applied to the device to create the electrolytic cells necessary to accelerate corrosion of the metallization or bond pads without heating the device. Most groups are tested to 1000 hours with some groups extended beyond to look for longer term effects. A typical test condition is: T_A = 85°C to 95°C, rh = 85% to 95%, and bias = 80% to 100% of the maximum rating. Common failure modes include parametric shifts, high leakage, and catastrophic. Common failure mechanisms include die corrosion or contaminants such as foreign material on or within the package materials, and poor package sealing. The purpose of the moisture–resistance test is to evaluate the moisture resistance of components under temperature/humidity conditions typical of tropical environments. A typical test condition is: T_A = –10°C to 65°C, and rh = 80% to 98%. Common failure modes include parametric shifts in leakage and mechanical failure. Common failure mechanisms include corrosion or contaminants on or within the package materials, and poor package sealing.

Autoclave

Autoclave is an environmental test that measures device resistance to moisture penetration and the resultant effects of galvanic corrosion. Conditions often employed during the test include 121°C, 100% relative humidity, and 15 psig. Corrosion of the die is the expected failure mechanism. Autoclave is a highly accelerated and destructive test. This test may be followed by other tests to further accelerate the corrosion failure mechanism. Common failure modes include parametric shift, high leakage, and catastrophic event. Common failure mechanisms include die corrosion, contaminants such as foreign material on or within the package materials, or poor package sealing.

Pressure/Temperature/Humidity Bias (PTHB)

This test is performed to accelerate the effects of moisture penetration with the dominant effect being corrosion. This test detects similar failure mechanisms as THB but at a greatly accelerated rate. Conditions usually employed during this test are a temperature of 125°C or greater, pressure of 15 psig or greater, humidity of 100%, and a bias level. The purpose of the moisture resistance test is to evaluate the moisture resistance of components under temperature/humidity conditions typical of tropical environments. A typical test condition is: T_A = –10°C to 65°C, and rh = 80% to 98%. Common failure modes include parametric shifts in leakage and mechanical failure. Common failure mechanisms include corrosion or contaminants on or within the package materials, and poor package sealing.

Cycled Temperature Humidity Bias

This test is used to examine the ability of devices to withstand the combined effects of temperature cycling, high humidity, and voltage. This test differs from a typical humidity test in its use of temperature cycling.

Power Temperature Cycling

This test is performed on semiconductor devices to determine the effects of alternate exposures to extremes of high and low temperature with operating voltages periodically applied and removed. This is more of a performance assessment than a reliability test. Test duration is normally for 1000 cycles with some tests being extended to look for longer–term effects. Common failure modes include parametric shifts and catastrophic events. Common failure mechanisms include wire bond, cracked or lifted die, and package failure.

Power Cycling

This test is performed at a constant ambient temperature with operating voltage(s) periodically applied and removed, producing a high operating junction temperature typically between 50°C and 150°C above ambient.

Low–Temperature Operating Life

This test is performed primarily to accelerate hot carrier injection effects in semiconductor devices by exposing them to room ambient or colder temperatures with the use of biased operating conditions. Threshold shifts or parametric changes are typically the basis for failure. The length of this test varies with temperature and bias conditions employed.

Mechanical Shock

This test is used to examine the ability of the device to withstand a sudden change in mechanical stress, typically due to abrupt changes in motion as seen in handling, transportation, or actual use. A typical test condition is: acceleration = 1500 g; orientation = Yl plane; t = 0.5 ms, and number of pulses = 5. Common failure modes include open, short, excessive leakage, and mechanical failure. Common failure mechanisms include die and wire bonds, cracked die, and package defects.

Constant Acceleration

This test is used to indicate structural or mechanical weaknesses in a device/packaging system by applying a severe mechanical stress. A typical test condition used is: stress level = 30 kg. Common failure modes include mechanical failure. Common failure mechanisms include poor package design.

Solder Heat

This test is used to examine the device's ability to withstand the temperatures seen in soldering over a more extended period as compared to the typical exposure levels seen in a production process. Electrical testing is the endpoint criterion for this stress. Common failure modes include parameter shifts, and mechanical failure. Common failure mechanisms include poor package design.

Lead Integrity

This test is used to examine the mechanical properties of a device's leads, welds, and seals. Various conditions can be employed and provided for: tensile loading, bending stresses, torque or twist, and peel stress. The failure is determined visually under 3x to l0x magnification.

Solderability

The purpose of this test is to measure the ability of device leads/terminals to be soldered after an extended period of storage (shelf life). Common failure modes include pin holes, dewetting, and nonwetting. Common failure mechanisms include poor plating and contaminated leads.

Variable Frequency Vibration

This test is used to examine the ability of the device to withstand deterioration due to mechanical resonance. A typical test condition is: peak acceleration = 20 g; frequency range = 20 Hz to 20 KHz. Common failure modes include open, short, excessive leakage, and mechanical failure. Common failure mechanisms include die and wire bonds, cracked die, and package defects.

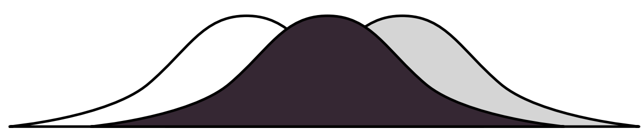
Statistical Process Control

Initial design improvement is one method that can be used to produce a superior product. Equally important to outgoing product quality is the ability to produce products that consistently conform to specifications. Process variability is the basic enemy of semiconductor manufacturing, since it leads to product variability. Statistical process control (SPC) replaces variability with predictability. Use of SPC methods assures the product will meet specific process requirements throughout the manufacturing cycle. The emphasis is on defect prevention, not detection. Predictability through SPC methods requires the manufacturing culture to focus on constant and permanent improvements.

Circuit performance is often dependent on the cumulative effect of component variability. Tightly controlled component distributions give the customer greater circuit predictability. These programs require improvements in cycle time and yield predictability achievable only through SPC techniques. The benefit derived from SPC helps the manufacturer meet the customer's expectations of higher quality and lower cost product.

Capability of 6σ means parametric distributions will be centered within the specification limits with a product distribution of ±6 σ about the mean. This capability, shown graphically in Figure 196, details the benefits in terms of yield and outgoing quality levels. This compares a centered distribution versus a 1.5 σ worst–case distribution shift.

To better understand SPC principles, brief explanations have been provided. These cover process capability, implementation, and use.



Standard Deviations From Mean

Distribution Centered	Distribution Shifted ±1.5
At ±3 σ 2700 ppm defective	66810 ppm defective
99.73% yield	93.32% yield
At ±4 σ 63 ppm defective	6210 ppm defective
99.98937% yield	99.379% yield
At ±5 σ 0.57 ppm defective	233 ppm defective
99.999943% yield	99.9767% yield
At ±6 σ 0.002 ppm defective	3.4 ppm defective
99.9999998% yield	99.99966% yield

Figure 196. AOQL and Yield from a Normal Distribution of Product with 6σ Capability

Process Capability

One goal of SPC is to ensure a process is capable. Process capability is the measurement of a process to produce products consistently to specification requirement. The purpose of a process capability study is to separate the inherent random variability from assignable causes. Once completed, steps are taken to identify and eliminate the most significant assignable causes. Random variability is generally present in the system and does not fluctuate. Sometimes, these are considered basic limitations associated with the machinery, materials, personnel skills, or manufacturing methods. Assignable–cause inconsistencies relate to time variations in yield, performance, or reliability.

Traditionally, assignable causes appear to be random due to the lack of close examination or analysis. Figure 197 shows the impact on predictability that assignable cause can have. Figure 198 shows the difference between process control and process capability.

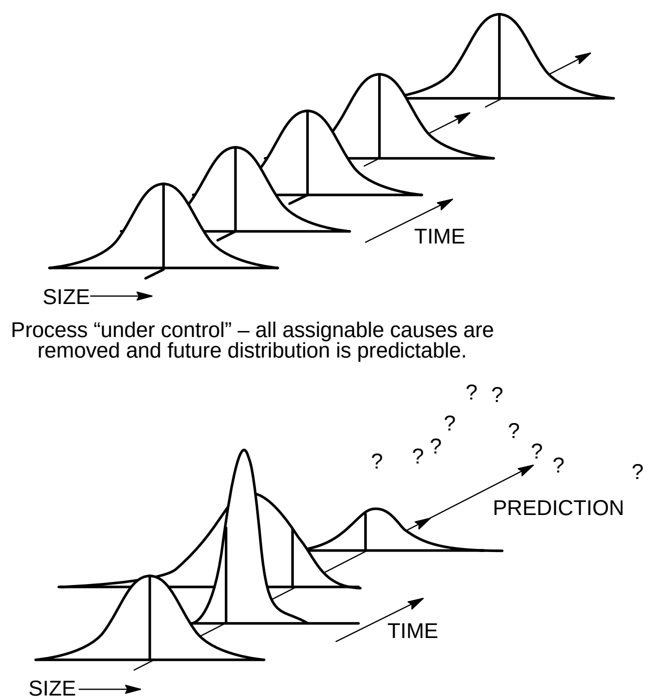


Figure 197. Impact of Assignable Causes on Process Predictability

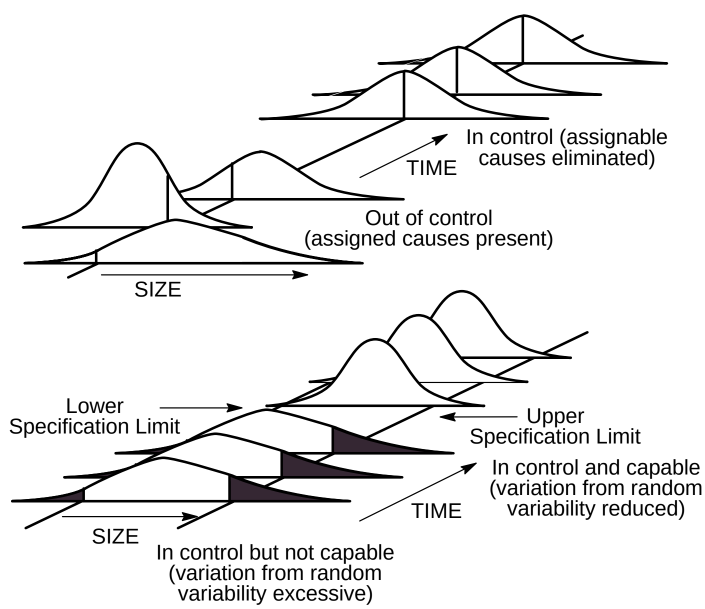


Figure 198. Difference Between Process Control and Process Capability

A process capability study involves taking periodic samples from the process under controlled conditions. The performance characteristics of these samples are charted against time. In time, assignable causes can be identified and engineered out. Careful documentation of the process is key to accurate diagnosis and successful removal of the assignable causes. Sometimes, the assignable causes will remain unclear, requiring prolonged experimentation.

Elements which measure process variation control and capability are Cp and Cpk, respectively. Cp is the specification width divided by the process width or Cp = (specification width) 16 σ. Cpk is the absolute value of the closest specification value to the mean, minus the mean ($\overline{X}$), divided by half the process width or Cpk = (closest specification) – $\overline{X}$ /3 σ.

For critical parameters, the process capability is typically acceptable with a Cpk = 1.33. The desired process capability is a Cpk = 2 and the ideal is a Cpk = 5. Cpk, by definition, shows where the current production process fits with relationship to the specification limits. Off–center distributions or excessive process variability will result in less than optimum conditions.

SPC Implementation and Use

Some parameters are sensitive to process variations while others remain constant for a given product line. Often, specific parameters are influenced when changes to other parameters occur. It is both impractical and unnecessary to monitor all parameters using SPC methods. Only critical parameters that are sensitive to process variability are chosen for SPC monitoring. The process steps affecting these critical parameters must be also identified. It is equally important to find a measurement in these process steps that correlates with product performance. This is called a critical process parameter.

Once the critical process parameters are selected, a sample plan must be determined. The samples used for measurement are organized into rational subgroups of approximately two to five pieces. The subgroup size should be such that variation among the samples within the subgroup remains small. All samples must come from the same source (for example, the same mold press operator, etc.). Subgroup data should be collected at appropriate time intervals to detect variations in the process. As the process begins to show improved stability, the interval may be increased. The data collected must be carefully documented and maintained for later correlation. Examples of common documentation entries would include operator, machine, time, settings, and product type.

Once the plan is established, data collection may begin. The data collected will generate $\overline{X}$ and R values that are plotted with respect to time. $\overline{X}$ refers to the mean of the values within a given subgroup, while R is the range or greatest value minus least value. When approximately 20 or more $\overline{X}$ and R values have been generated, the average of these values is computed. The values of X and R are used to create the process control chart. Control charts are the primary SPC tool used to signal a problem.

Once the variability is identified, the cause of the variability must be determined. Normally, only a few factors have a significant impact on the total variability of the process. Most techniques may be employed to identify the primary assignable cause(s). Out–of–control conditions may be correlated to documented process changes. The product may be analyzed in detail using best– versus worst–part comparisons or product analysis lab equipment. Multivariance analysis can be used to determine the family of variation (positional, critical, or temporal). Lastly, experiments may be run to test theoretical or factorial analysis. Whatever method is used, assignable causes must be identified and eliminated in the most expeditious manner possible.

After assignable causes have been eliminated, new control limits are calculated to provide a more challenging variability criteria for the process. As yields and variability improve, it may become more difficult to detect improvements because they become much smaller. When all assignable causes have been eliminated and the points remain within control limits, the process is said to be in a state of control.

Chapter 13

Cooling Principles

Cooling Principles

In many high current rectifier diode applications, the losses are large enough that a heat exchanger is required in order to prevent the junction temperature from exceeding its rated limit and running the risk of failure. This chapter provides background information for a designer to investigate the thermal management considerations for rectifiers. Design charts and discussions of basic principles are presented. The focus is on the thermal analysis of heat exchangers. However, tradeoffs involving device performance and reliability, the environmental usage conditions, the accessibility to a coolant, and cost must also be considered as part of the design process.

Thermal Resistance Concepts

Heat is transferred by conduction (heat travels through a material), convection (heat is transferred by physical motion of a fluid), and radiation (heat is transferred by electromagnetic wave propagation). Semiconductors depend on conduction as a means of transferring heat from the device junction to the external surface of the package. Heat transfer from the package to the external environment depends on the method of interconnection of the package (i.e., leaded or leadless connection to a heat exchanger or printed wiring board) and the dominant mode of heat transfer (i.e., convection typically dominates when forced air or liquid cooling is used, while radiation may dominate in vacuum or high altitude applications).

The basic equation for heat transfer is:

$$q = hA\Delta T \qquad (13.1)$$

where q = rate of heat transfer (W),

h = heat transfer coefficient or thermal transmittance ($W/in^2/°C$),

A = cross–sectional area normal to the path of heat flow (in^2),

ΔT = temperature difference between regions of heat transfer (°C).

Electrical engineers find it easier to work in terms of thermal resistance. From Equation 13.1, die thermal resistance R_θ is:

$$R_\theta = \frac{\Delta T}{q} = \frac{1}{hA} \qquad (13.2)$$

where the coefficient h depends upon the heat transfer mechanism.

When determining cooling capability, it is simpler to use h rather than R because cooling due to convection and radiation may be handled as two thermal paths in parallel.

An analogy between Equation 13.2 and Ohm's Law is often made to form models of heat flow. Note that ΔT could be thought of as a voltage, thermal resistance corresponds to electrical resistance (R), and power (q) is analogous to current (T). This gives rise to the basic thermal resistance model for a semiconductor mounted to a heatsink indicated by Figure 199. Although a case–mounted part is shown, the model is generally applicable. For leadless surface–mounted parts, the dominant path for heat transfer is through the package case. With lead–mounted rectifiers, most of the heat leaves the junction via conduction through the leads. Thermal models for lead–mounted parts are discussed in Chapter 2.

The equivalent electrical circuit may be analyzed using Kirchoff's Law and the following equation results:

$$T_J = P_D\,(R_{\theta JC} + R_{\theta CS} + R_{\theta SA}) + T_A \qquad (13.3)$$

where $R_{\theta JA}$ = total thermal resistance junction to ambient ($R_{\theta JA} + R_{\theta CS} + R_{\theta SA}$)

$R_{\theta JC}$ = semiconductor thermal resistance (junction to case),

$R_{\theta CS}$ = interface thermal resistance (case to heatsink),

$R_{\theta SA}$ = heatsink thermal resistance (heatsink to ambient).

With rectifiers having significant reverse power losses, junction–to–ambient thermal resistance must be considered early in the design. The thermal characteristics of the semiconductor are described in Chapter 2.

The interface case–to–sink thermal resistance ($R_{\theta CS}$) resulting from imperfect mating surfaces and use of insulators may be appreciable in high–current applications. Factors involved in minimizing $R_{\theta CS}$ are discussed in the next chapter.

This chapter is devoted to explaining the variables affecting the heatsink to ambient thermal resistance $R_{\theta SA}$. It is assumed that the design problem has been reduced to one where the total power dissipation of the rectifier and the required heatsink temperature limit are known. Knowing the upper ambient temperature permits Equation 13.3 to be solved for the limit $R_{\theta SA}$.

The thermal resistance of the heat exchanger, or heatsink, is an important part of the thermal design, especially for stud– or base–mounted rectifiers. Heatsinks may transfer heat to the ambient by natural convection and radiation, or forced air or liquid cooling may be employed. As a rough approximation for silicon rectifiers, for dissipating power levels over 100 W, forced air cooling tends to be most economical in terms of space and cost; for power levels over 2000 W, liquid cooling is typically required. Schottky rectifiers, because of their lower maximum junction temperature limit, may require forced air or liquid cooling at lower power levels. However, the current levels are roughly the same because of the lower forward voltage of the Schottky diode. In terms of thermal resistance, values as low as 1°C/W can be achieved with large volume free convection heatsinks. Use of air flow can produce values as low as 0.05°C/W, while use of liquid cooling can closely approximate the infinite heatsink.

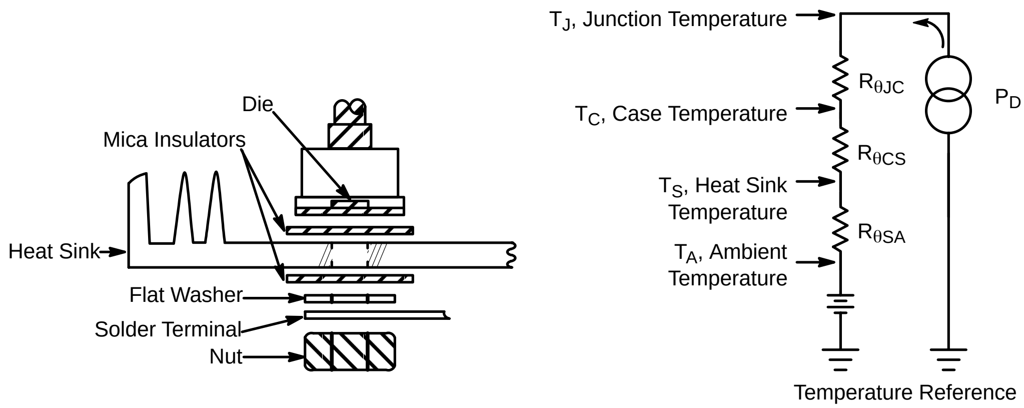


Figure 199. Basic Thermal Resistance Model Showing Thermal to Electrical Analogy for a Semiconductor

In the following sections, factors of importance in natural convection cooling, forced–air cooling, and liquid cooling are discussed. The intent is to acquaint the electrical engineer with the basic principles involved. Reference [1] provides a place to begin further study and has an extensive list of references.

Conduction

Conduction is a process of heat transfer in which heat energy is passed from one molecule to the next, while the actual molecules involved in the transfer remain in their original positions. In an opaque solid, heat conduction is the only means of heat transfer, where heat flows from the hot to the colder areas of the solid. Conduction can occur in liquids and gases, but the amount of heat transferred is generally smaller for the same geometry.

The thermal resistance attributed to conduction, $R_{\theta\,(cond)}$, is often modeled in terms of the one–dimensional equation:

$$R_\theta(\text{cond}) = X / k_\theta A \qquad (13.4)$$

where $R_{\theta\,(cond)}$ = thermal resistance (°C/W),
X = length of thermal path (in),
k_θ = thermal conductivity (W/in°C),
A = area normal to the thermal path (in^2).

(The heat transfer coefficient h is k_θ/X.) The coefficient of thermal conductivity is a material property, which can range over five orders of magnitude. Thermal conductivity for a number of materials used in electronics is given in Table 26.

Heatsink design often involves cost–effectively minimizing conduction resistance. In large area plates, conduction resistance may be significant and may play a key part in determining a design in which the volume and weight of the heatsink must also be minimized.

Convection

Convection is the transfer of heat by the physical motion or mixing action of a fluid, such as air or water. When the movement is due entirely to temperature differences within the fluid, which results in differences in density, the mechanism is called natural convection. Note that in a gravity field, heated air expands, becomes less dense and therefore lighter than the surrounding air, and rises. This in turn creates the “natural” movement of air. When the motion of the fluid is produced by mechanical methods, such as fans and pumps, the mechanism is called forced convection.

Table 26. Thermal Conductivity of Various Materials at 27°C

Materials	Conductivity W/in–°C
Silver	10.7
Copper	9.75
Gold	7.86
Aluminum, pure	5.75
Beryllia, 99.5% pure	6.20
Aluminum, 68S	5.34
Beryllia, 95%	4.13
Magnesium	4.16
Aluminum 6061–T6	4.13
Red brass	2.92
Yellow brass	2.51
Beryllium cooper	2.19
Pure iron	1.99
Phosphor bronze	1.36
Soft steel	1.23
Monel	0.94
Hard steel	0.58
Thermal compound	0.018
Mica	0.015
Mylar film	0.0042
Air	0.00083

Natural Convection

The following equation applies for the natural convection of vertical plates (heatsink) suspended in free air at ground level:

$$h_C = 2.21 \times 10^{-3}\left(\frac{T_S - T_A}{L}\right)^{0.25} \qquad (13.5)$$

where h_c = convection heat transfer coefficient (W/in^2°C),

L = height of the heatsink (in),

T_S = surface temperature of the heatsink (°C).

Figure 200 is a plot of Equation 13.5. The average natural convection coefficient h_c defines the thermal characteristics of the air film that encapsulates the plate. The constant is a function of air density, temperature of ambient air in contact with the plate, and orientation and shape of the heatsink. The value of 2.21 x $l0^{-3}$ applies for standard pressure and temperature at sea level for rectangular plates suspended freely in still air. For other conditions, the “constant” must be altered [2].

Air density is a significant factor in convection cooling. Since density varies with altitude, a correction factor is indicated on Figure 201. Note that a 10% loss in effectiveness occurs at about 5000 ft. In space, convection is zero. Values for h_c must be multiplied by the number obtained from Figure 201 to obtain the proper value.

Plate shape and orientation also affect convection cooling. Table 27 indicates procedures for obtaining the significant dimension L and a correction factor F_c for rectangular, circular, and cylindrical surfaces in either a horizontal or vertical position. For horizontal mounting, the sum of the convection thermal transmittance from the top and bottom surfaces must be found. When both surfaces are exposed to air, the overall thermal resistance is about 25% higher than would be the case of a vertical plate of the same significant dimension L. For an enclosed chassis, the contribution of the bottom surface of the upper plate is negligible.

As an example of the use of the data, the convection thermal resistance of a 4 in. x 6 in. plate will be found (total surface area = 4 x 6 x 2 = 48 in^2). Assume the surface temperature is 120°C and the ambient is 60°C. For mounting with the 4 in. side in a vertical plane (the most efficient mounting), read h_c = 4.4 x 10^3 W/in^2°C from Figure 200. Using Equation 13.2, $R_{\theta(conv)}$ = 1/(4.4)(10^{-3})(48) = 4.74°C/W. If the plate were mounted with the 6 in. side in a vertical plane, h_c = 4 x 10^{-3} W/in^2°C resulting in a 10% increase in $R_{\theta(conv)}$.

For horizontal mounting, a new value of L must be found from the relations in Table 27. Dimension L calculates to be 2.4 in. resulting in h_c = 5 x $l0^{-3}$ W/in^2°C at T_S–T_A = 60°C. The area of each side is 24 in^2. Using the correction factors given in Table 27, h_c(top) = (0.9)(5) 10^{-3} = 4.5 x 10^{-3} The contribution of the bottom surface, if it is exposed, is 0.45(5)10^{-3} = 2.25 x $l0^{-3}$ W/in^2°C. The total value of $R_{\theta(conv)}$ = 1/(4.5 + 2.25)(10^{-3})(24) = 6.15°C/W. Thus, the horizontal mounting result is about 30% worse than the most favorable vertical position.

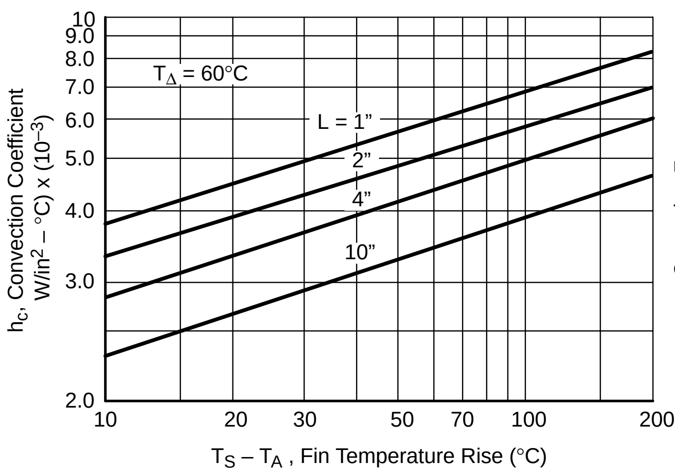


Figure 200. Free Convection Coefficient for Rectangular Plates of Various Heights, *L*

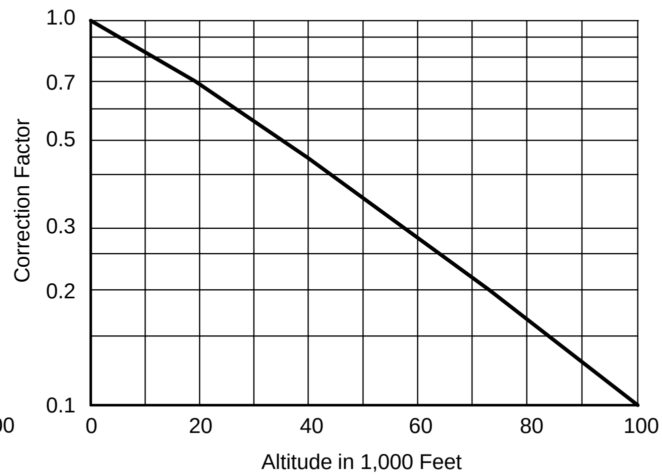


Figure 201. Altitude Correction Factor for Free Convection Heat Transfer

Table 27. Significant Dimension L and Correction Factor F_c for Convection Thermal Resistance

Surface	Significant Dimension L		Correction Factor F_c	
	Position	L	Position	F_c
Rectangular Plane	Vertical	Height (max 2 ft)	Vertical Plane	1.0
	Horizontal	Length x Width/(Length + Width)	Horizontal Plane, facing downward	0.45
Circular Plane	Vertical	π/1 x Diameter	Horizontal Plane, facing upward	0.9
Cylinder	Horizontal	Diameter	Horizontal	0.82
	Vertical	Height (max 2 ft)	Vertical	0.9

Forced Convection

In forced convection a fluid is moved by or through the heatsink by an external source, such as a fan or a pump. Because the fluid flow rate may be quite high, forced convection is capable of reducing the heatsink thermal resistance to extremely low values.

Forced–air convection produces a drastic reduction in thermal resistance of any heatsink as compared to natural convection due to a “scouring” effect upon the insulating boundary layer of dead air at the fin surface. Natural convection cooling is primarily dependent upon the volume and material of the heatsink, while forced convection cooling can be related to the airflow, fin spacing, and the degree of coupling to the air stream.

The heat transfer coefficient for the usual case of turbulent air flow parallel to a plate may be expressed as:

$$h_f = 6.63 \times 10^{-3} (V^3/L)^{0.25} \qquad (13.6)$$

where h_f = forced convection heat transfer coefficient ($W/in^2 - °C$),

V = free stream velocity (linear feet per minute),

L = length of the plate (in).

The constant is affected by air density and temperature. The value given is based upon operation at sea level at an average air temperature over the heat exchanger of 600°C. As a design aid, values have been calculated from Equation 13.6 and are plotted in Figure 202. Note that a short dimension in the flow direction increases the coefficient significantly. The curves stop at points where the flow is becoming laminar and consequently Equation 13.6 does not apply. In the laminar region, h_f is higher than an extrapolation of the curves would indicate. Note that Equation 13.6 should not be used for plates with L less than 4 in., as the flow is laminar when it first encounters the plate. Short plates do not permit development of turbulent flow.

As an example, the forced convection thermal resistance will be found for the 4 in. x 6 in. plate (surface area = 48 in^2) used in the previous example. Assume that the 4 in. side is parallel to the flow direction and the air flow rate is 2000 LFM. Reading from Figure 202, $h_f \approx 1.4$ $W/in^2 °C$. Therefore, R_θ = 1/(l/4)(48) = 0.015°C/W. This value is about 300 times better than that obtained with natural convection. However, improvements this large are not achieved in practice because of material thermal resistance losses.

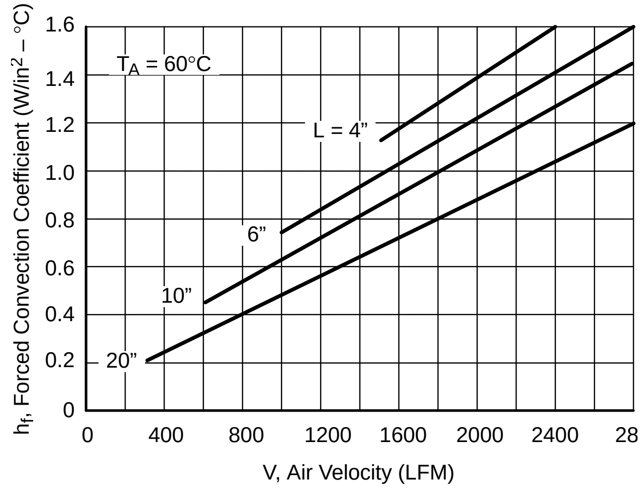


Figure 202. Heat Transfer Coefficient for Forced Air Convection

Water or liquid cooling is similar in principle to that of forced air. The heat transfer coefficient is related to velocity in a manner similar to that described by Equation 13.6. Normally, the liquid cooled system is purchased as a unit whose characteristics are provided by the manufacturer.

Radiation

The third process by which heat can be transferred is radiation. The ability of a body to radiate thermal energy at any particular wavelength is determined by the body temperature and surface characteristics. An ideal radiator is called a blackbody which, by definition, radiates the maximum amount of energy at any wavelength. The ratio of energy emitted by any surface to that of a blackbody at the same temperature is called the emissivity ε.

Incident radiation on a surface is partially absorbed, reflected, and, in some cases, transmitted through the surface. However, materials used for heatsinks are opaque and do not transmit radiation. The ability of a surface to absorb the total incident radiation on it is defined by the term absorptivity a. As with the emissivity, the absorptivity of a surface is dependent upon the source temperature (i.e., wavelength of incident radiation) and the receiving body surface characteristics.

Many materials exhibit the property of having a = e when the absorber and the emitter are at temperatures within 500°C of each other. These materials are called gray bodies and emissivity values for some common materials are listed in Table 28.

The basic equation for the thermal transmittance or heat transfer coefficient when radiation is taking place between two gray bodies is:

$$h_r = 3.68 \times 10^{-3}\epsilon\left[\left(\frac{T_S}{100}\right)^4 - \left(\frac{T_A}{100}\right)^4\right] \Big/ (T_S - T_A) \qquad (13.7)$$

Table 28. Typical Emissivities of Common Surfaces at 25°C

Surface	Emissivity
Aluminum, Anodized	0.7–0.9
Alodine on Aluminum	0.15
Aluminum, Polished	0.05
Copper, Polished	0.07
Copper, Oxidized	0.70
Rolled Sheet Steel	0.66
Air Drying Enamel	0.85–0.91
Oil Paints	0.92–0.96
Varnish	0.89–0.93

where h_r = radiation heat transfer coefficient (W/in^2°C)
ε = emissivity,
T_S= surface temperature (°K),
T_A = ambient temperature (°K).

As with convection, the thermal resistance is inversely proportional to area. Since temperature affects the results in a complex manner, Equation 13.7 is plotted in Figure 203 in terms of 0°C. The strong effect of temperature upon the radiation coefficient is apparent. As with convection, a large temperature drop can occur in large area plates due to resistance losses, whereby the extremities of the heatsink are not as effective as the center.

If radiation is obstructed by nearby objects, the “constant” in Equation 13.7 will need to be modified by a factor (f_r) which is smaller than unity. If the object is at the same temperature as the fin, an adjustment can be made by considering the geometry of the radiation pattern. It is generally satisfactory to consider an unobstructed fin’s radiation as originating at the center of the fin and being spherical (hemispherical on each side of the fin). An obstruction will interrupt radiation, or subtract a sector from the sphere; f_r is approximately the ratio of the solid angle remaining in the obstructed sphere to the solid angle (4 π steradians) of the complete sphere.

The fin should be shielded from bodies of higher temperature. Otherwise, the fin will be heated by radiation instead of being cooled.

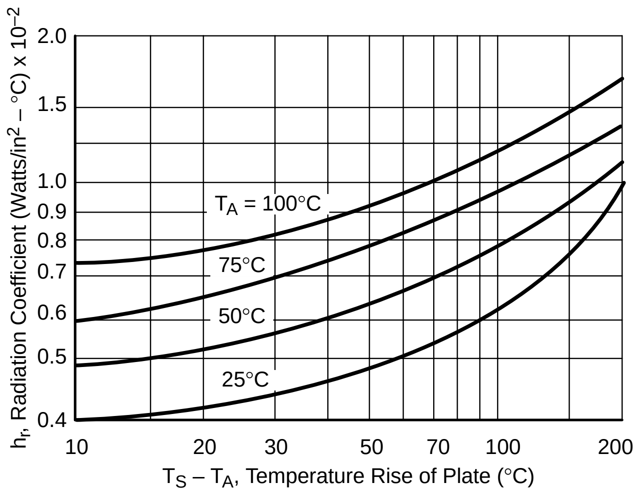


Figure 203. Radiation Coefficient for a Black Body having an Unobstructed Pattern

Equation 13.7 also assumes that the radiating body is small compared to the enclosing body. Other configurations require a modification of the emissivity value. For the case of two large parallel plates or for the case where the enclosed body is large compared to the enclosing body, the effective emissivity e_f is given by:

$$\epsilon_f = \frac{1}{\frac{1}{\epsilon_1} + \frac{1}{\epsilon_2} - 1} \qquad (13.8)$$

where: ε_1 and ε_2 are the emissivities of the two surfaces. Other situations require an in–depth analysis of the problem.

As an example of the use of the data for radiation thermal resistance, R(rad) for the 4 in. x 6 in. plate previously used will be found. From Figure 203, at T_A = 60°C and T_S = 120°C, h_t = 0.7 x 102 W/in^2°C. If the plate were painted (preferably black), ε = 0.9 and $R_{\theta(rad)}$ = 1/(0.9)(0.7)(10^{-2})(48) = 3.3°C/W. This value is less than the convection value because of the high emissivity and high surface temperature allowed.

Surface emissivity is unimportant when forced air or liquid cooling is used, because the effect of radiation is negligible. However, for natural convection cooling, radiation plays a significant roll and surfaces having high emissivity should be used. As Table 28 shows, polished surfaces should be avoided and anodized aluminum and painted surfaces are preferable. Tests on natural convection heatsinks which are made commercially reveal a decrease in total thermal resistance of approximately 25% at high ambient temperatures when a dull black finish is used as opposed to a bright mill finish.

Fin Efficiency

Because of resistance losses in the fin material, both conduct on and radiation cooling efficiencies decrease with distance from the heat source. Fin efficiency is defined as the ratio of total heat dissipated by an identically shaped fin of infinitely conducting material under the same conditions. Knowing the efficiency, (η). the fin dissipation, (q), may be calculated from:

$$q = \eta A h_t (t_S - t_A) \qquad (13.9)$$

where A = fin surface area,

h_t = the total heat transfer coefficient,

T_S = the temperature of the heatsink at the heat source,

T_A = the temperature of the ambient.

Similarly the fin thermal resistance R_θ is:

$$R_{\theta SA} = \frac{1}{\eta A h_t} \qquad (13.10)$$

If fins are made of relatively thick, highly conductive material, fin efficiency is close to 100% and the total fin thermal conductance is the sum of the convection and radiation components. At the other extreme, if the fin is quite thin and/or the material has low conductivity, the fin extremities are ineffective because of high resistance losses.

Both extremes are wasteful of material and costly but the latter extreme is encountered when a chassis or other structure is used as a cooling fin for semiconductors. The range between these extremes is rather narrow.

The analysis of fin efficiency is more manageable when performed using a circular fin, and when heat transfer by means of convection and radiation from the fin edge is neglected. The more commonly used geometries are handled by converting their dimension to an equivalent radius using the formulas in Table 29. The fin thickness (S) is generally a negligible factor in the equivalent radius.

An important factor in fin efficiency is the heat input radius, r_i. Figure 204 shows how it is found for the popular stud and diamond base packages. It is essentially the average value of the dimensions of the heat input source.

By analyzing the temperature drop and heat dissipated by a volume element in an annular ring, a parameter is discovered which is a characteristic of the material and the heat transfer coefficient. This parameter, which plays an important role in fin efficiency, is called the “natural radius” of the fin (R) and is given by:

$$R = \sqrt{\frac{k_\theta S}{2h_t}} \qquad (13.11)$$

where k_θ = the thermal conductivity of the fin material,

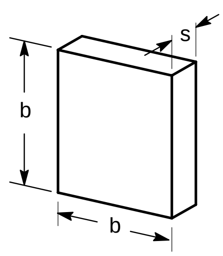
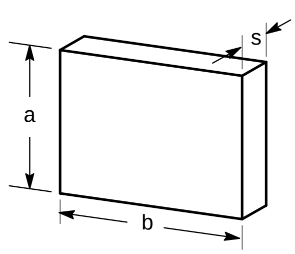
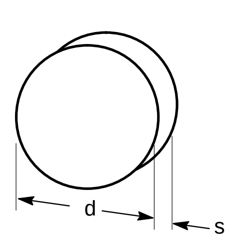
S = fin thickness

h_t = total heat transfer coefficient of the fin.

For convenience in problem solving, Equation 13.11 is plotted in Figure 205 for 63S aluminum plates of standard thickness. Since the term $k_\theta S$ is a constant for any one of the curves of Figure 205, the data may be used for heatsinks of other material by considering the thermal conductivity data of Table 26. For example, since copper has almost twice the conductivity of 63S aluminum, a copper plate of 1/32 in. has the same relationship of h_t to S as does a 1/16 in. aluminum plate.

Further analysis yields the equations necessary to plot Figure 206 and the fact that an optimum fin design is achieved when the fin efficiency is 50%. An optimum fin is taken as one which has minimum volume and therefore minimum material which generally results in minimum cost. For the optimum fin, the outer radius is approximately equal to the natural radius (when the fin outer radius is large compared to the heat input radius).

Table 29. Equivalent Outer Radius (r_0) for Various Fin Geometries

		
$\sqrt{\frac{b^2 + 2bs}{\pi}}$	$\sqrt{\frac{ab + (a+b)s}{\pi}}$ (Satisfactory for $b \leq 2a$)	$\frac{d}{2}\sqrt{1 + \frac{2s}{d}}$

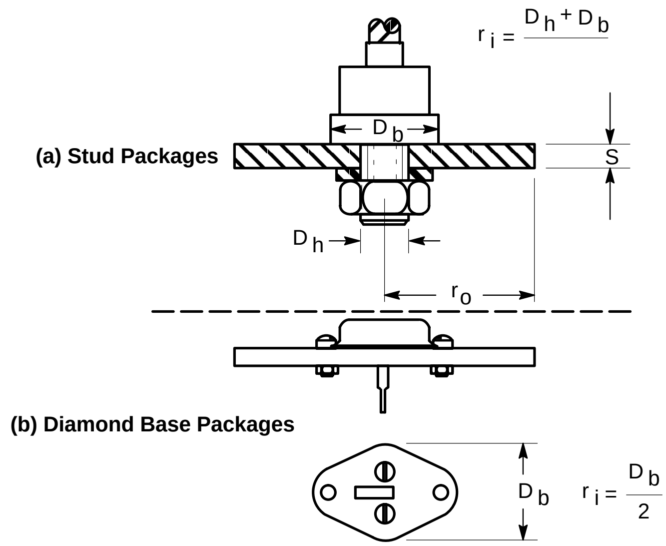


Figure 204. Relation of Semiconductor Device Dimensions to Heat Input Radius r_i

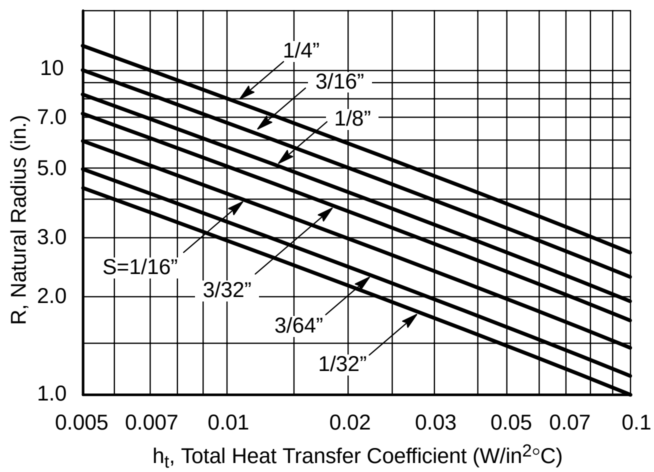


Figure 205. Effect of Fin Thickness and Heat Transfer Coefficient upon the Natural Radius of Type 63S Aluminum Fins

To illustrate the use of the data, the efficiency of the 4 in. x 6 in. plate used in the previous examples will be discussed. Assume a DO–5 package is attached which has an effective heat input radius close to 0.22 in. The effective outer radius r0 from Table 29 is 24/π (neglecting S) or 2.78 in. Therefore, r_0/r_i = 12.6. For an optimum fin (h = 50%) R/r_0 12.6 from Figure 206. Therefore R = 2.78 in. The fin thickness is determined from Equation 13.11 or Figure 205. From the previous two examples, $h_t = (0.63 + 0.44)\ 10^{-2} = 0.0107$. Therefore, S maybe less than 1/32 in. for an aluminum fin. Overall thermal resistance is found from Equation 13.10 and is 3.9°C/W for this example.

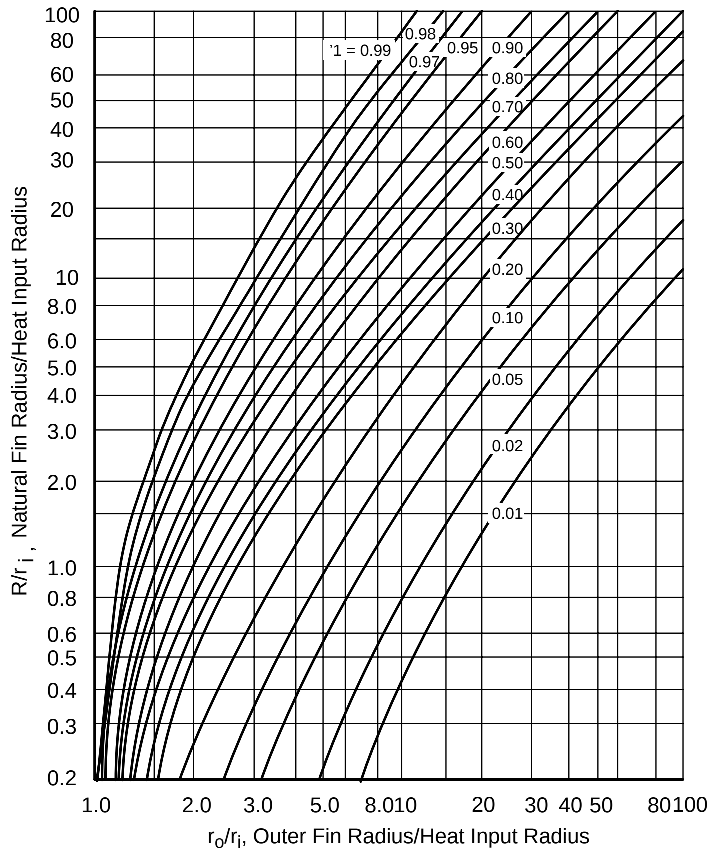


Figure 206. Fin Efficiency η as a Function of R/r_i and r_o/r_i

An improvement in fin efficiency or a reduction in fin size for a desired thermal resistance is achieved by increasing the radius of the heat source. In practice, this can be achieved by brazing copper material to the heatsink. Suppose that 2 in. diameter discs are brazed to each side of the plate in the above example and that the thermal resistance of the discs is negligible. Now r_0/r_i = 2.78 and R/r_i = 2.78 also. From Figure 206, the fin efficiency is increased to 80%. The advantage of a large highly conductive semiconductor package over a smaller one is thus also apparent.

Application and Characteristics of Heatsinks

The material in this section provides general practical information necessary to select and use heatsinks in free and forced convection. In most cases, a commercial heatsink is used in manufactured equipment. The most important characteristics of heatsinks are specified by the manufacturer. By using the basic principles of heat transfer, the behavior of heatsinks under a variety of conditions may be predicted. Furthermore, the data for Figures 201 through 206 may be used to design single–fin heatsinks and to obtain the thermal resistance of a metal chassis or a PCB.

Heatsinks for Free Convection Cooling

Heatsinks can be grouped into three categories:

1. Flat vertical–finned types. Normally aluminum extrusions with or without an anodized black finish, they are unexcelled for natural convection cooling and provide reasonable thermal resistance at moderate air–flow rates for forced convection.
2. Cylindrical or radial vertical–finned types. Normally cast aluminum with an anodized black finish, they are used when maximum cooling in minimum lateral displacement is required.
3. Cylindrical horizontal–finned types. Normally fabricated from sheet–metal rings with a painted black matte finish, they are used in confined spaces for maximum cooling in minimum vertical displacement but are less efficient than the other two types.

The thermal resistance can be related fairly well to the surface area and the volume of the heatsink as illustrated by Figures 207 and 208. The data of most interest is for heatsinks of categories 1 and 2; the category 3 types do not fit into a pattern and are of little interest for power rectifiers. The steady–state characteristics typical of a category 1 heatsink are shown in Figure 209. The slope of the curve varies from 3°C/W at low power levels to about 2°C/W at high levels. The improvement in thermal resistance results from the convection and radiation components becoming more effective as temperature increases. The transient characteristics are shown in Figure 210. Note that the temperature below 2 minutes is proportional to the square root of time, a characteristic of one dimensional heat flow. The time to reach equilibrium is dependent upon the mass of the heatsink and requires about 20 minutes. Therefore, patience is required when evaluating the overall steady state thermal behavior of a system. Note that at times much less than the time for thermal equilibrium, the effective thermal resistance can be quite low, which can be used to advantage when the power diode is used intermittently.

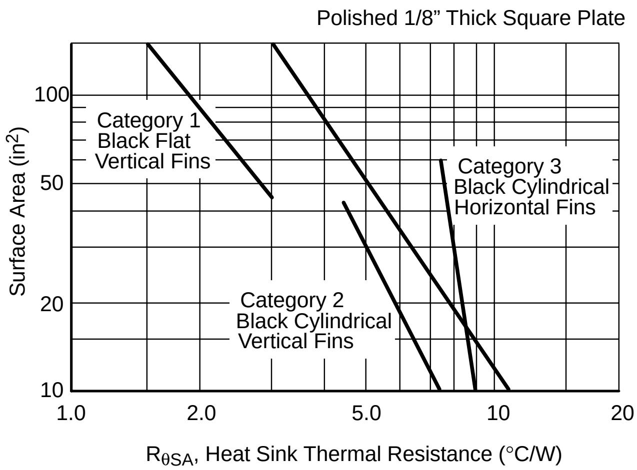


Figure 207. Relation of Surface Area to Thermal Resistance for Categories of Commercial Aluminum Heatsinks as Compared to an Aluminum Plate

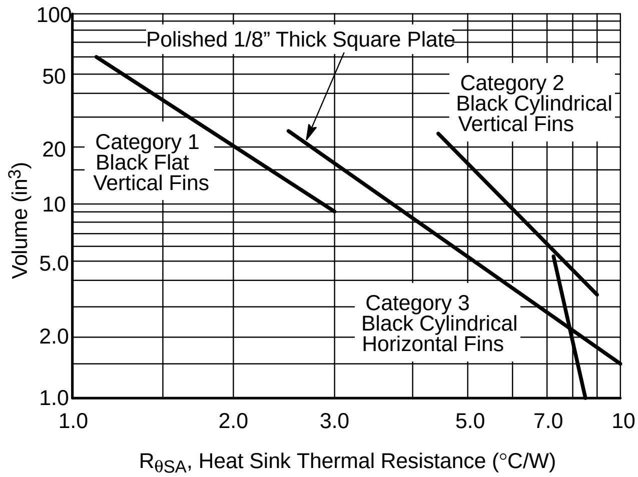


Figure 208. Relation of Volume to Thermal Resistance for Categories of Aluminum Heatsinks as Compared to an Aluminum Plate

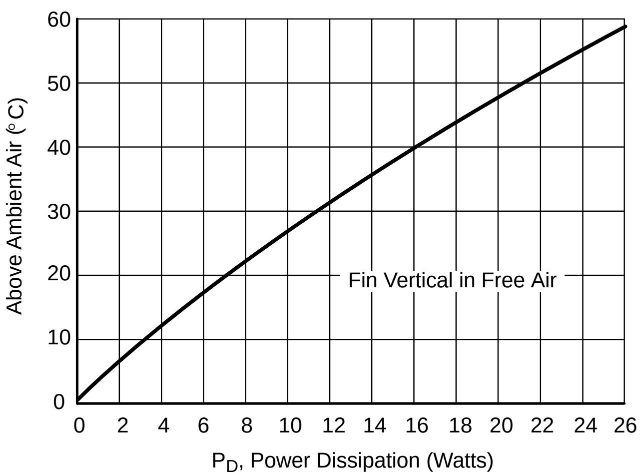


Figure 209. Typical Category 1 Heatsink Natural Correction Thermal Characteristics

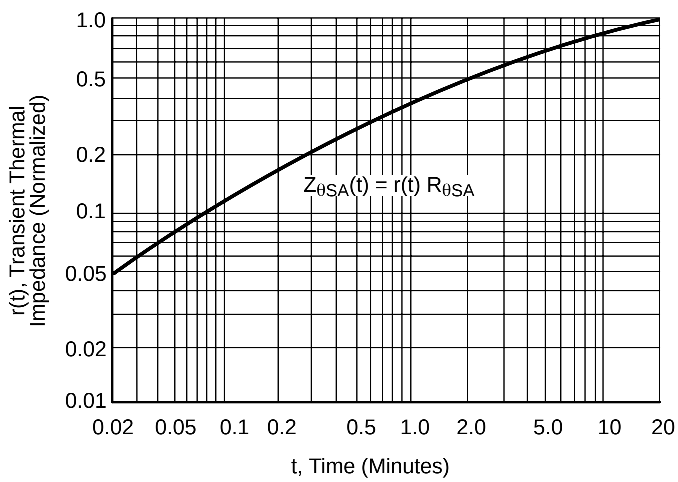


Figure 210. Thermal Response of Typical Category 1 Heatsink

Heatsinks for Forced Air Cooling

A natural reaction, when a designer finds that a semiconductor is running too hot when mounted to the largest heatsink permissible in the system, is to add a fan or blower. Although results may seem satisfactory, a more compact and less expensive cooling system generally results if a heat exchanger designed for use with forced air is used. Improvement in $R_{\theta SA}$ due to forced air is illustrated in Figure 211. The asymptotic behavior at high flow rates is typical of forced air cooling and indicates that the heatsink is becoming conduction limited.

When forced air cooling is employed, an interlock is generally necessary to prevent catastrophic system failure in the event of a blower malfunction. The use of an air–switch, comprised of a moving vane in the air flow mechanically coupled to a microswitch, can be used to interlock the electrical system.

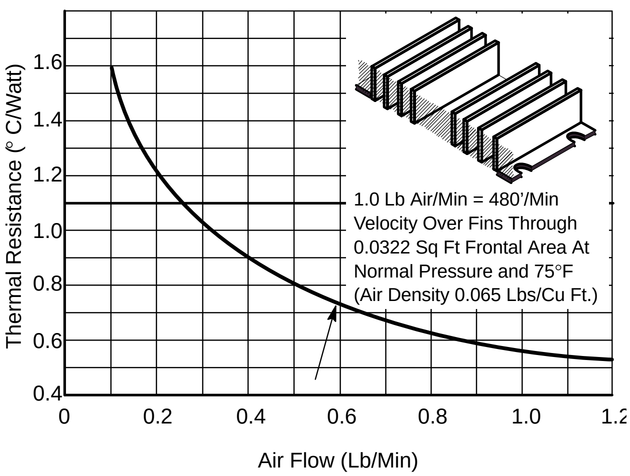


Figure 211. Performance Under Forced Airflow of Typical Category 1 Natural Convection Heatsink

The basic air flow rate equation is:

$$\Delta T = 1.76\, P_D / VA \qquad (13.12)$$

where ΔT = the change in air temperature (°C),
P_D = the power dissipation (W),
V = air velocity in linear feet per minute (LFM),
A = area of duct or chamber (ft^2).

For example, if a flow rate of 500 LFM is required in a one square foot duct and the power to be dissipated is 1000 W, the air temperature will increase by 3.5°C. The product VA is the volume of air flow required in cubic feet per minute (CFM).

Most efficient use of a given air flow will be achieved by locating components demanding minimum temperature rise (for example, semiconductors) closer to the inlet end of the cooling column and locating those elements for which maximum temperature rise is permitted (for example, power resistors) at the exhaust end.

Determination of the required air flow must also take into account the location of the airmover (fan or blower). If the airmover is located at the intake, its own heat loss must be added to the power which the system is required to dissipate. In this location, however, the ambient temperature which the blower experiences will be relatively low. At the exhaust, the airmover operates in a higher ambient temperature but its own power loss does not raise the ambient air of the assembly and is, therefore, generally preferable.

Pressure Drop

An important aspect of forced air cooling is that of pressure drop. There is always a frictional resistance to fluid flow which is a function of geometry, velocity, and properties of the fluid. The resistance causes a back pressure, often called pressure drop, ΔP.

For a passage of given geometry, pressure drop will increase with the square of air flow. With constant air flow, changes in the geometry that either restrict area or interfere with flow will result in an increase in pressure drop.

Airmovers deliver a decreasing amount of airflow as the resistance to flow or pressure rise increases. Maximum flow occurs at zero pressure rise and vice versa. Thus, for an airmoving system, a unique operational point is determined by the combination of the pressure drop and flow rate characteristics of the airmover and the heat exchanger as illustrated in Figure 212.

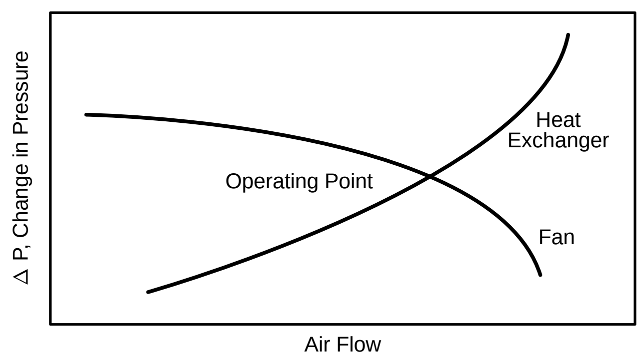


Figure 212. System Operating Point. The Operational Point of a Cooling System of Given Geometry and a Given Fan Occurs at the Intersection of the System's Pressure Drop vs. Air Flow, and of the Fan's Pressure Rise vs. Air Flow

Except for long straight ducts and certain bends and transition pieces, pressure drop is difficult or impossible to calculate. It is generally simpler to mount any type of airmover in the system and measure the resulting flow and pressure. Even if only one point is obtained, a system characteristic or impedance curve may be derived by assuming the impedance curve is quadratic. Knowing the CFM required, the resulting pressure is read from the impedance graph and an airmover meeting these requirements may be selected.

Note that an increase in flow from 50 to 100 CFM on the curve would increase the pressure requirements from 1.0 in. to 4.0 in., which is a large increase. At a certain point, if more air is required, the designer is well advised to open air passages in the system so that the impedance is lowered, rather than increasing the pressure capabilities of the airmover.

For a given flow rate and fixed volume, any change in geometry that increases the heat transfer will increase pressure drop causing the operational point on Figure 212 to move to the left. Therefore, for every system–and–fan combination, there is an optimum fin geometry. If volume is critical, it is best to select a coder–fan combination from a manufacturer. Often, semiconductor manufacturers offer rectifier heatsink assemblies designed for forced air service such as that shown in Figure 213.

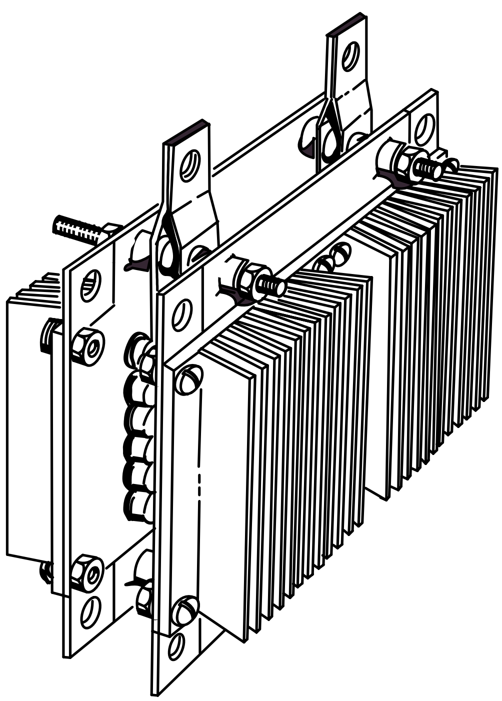


Figure 213. A 650 A Rectifier for 3–Phase Service Having Integral Heatsink Designed for Forced Air Systems

Fan Laws

1. Speed: The following fan laws express the manner in which the various quantities, in any fixed system, vary with speed of the air moving devices:
   a. CEM varies directly with rpm,
   b. SP (static pressure) varies as the square of the rpm,
   c. HP (horsepower) varies as the cube of the rpm.

Note that if the impeller speed is doubled, the CFM is doubled, the static pressure is multiplied by four and the horsepower required is multiplied by eight.

2. Density: The performance rating of any impeller acting at a constant speed in a fixed system will be altered by a change in air density in the following manner:
   a. CFM remains constant,
   b. SP varies directly with air density.

Density is as important as velocity in forced convection cooling. Since density decreases with altitude, an airmover which has just sufficient capacity at sea level is inadequate at 50,000 ft. In order to provide an airmover that will function efficiently at all altitudes, special slip motors are available which operate at higher speeds as the load decreases. Consequently, the motor maintains a constant mass flow rate in terms of pounds per minute, in spite of the changes in density caused by altitude.

Liquid Cooling

For very high current requirements, liquid cooling is required. Heat exchanger manufacturers' catalogs should be consulted for selections of suitable systems. The general principles outlined under air cooling also apply for liquid cooling. That is, Equation 13.7 applies, the pressure drop, flow rate relationship is quadratic, and the fan laws also apply. Typical performance of a liquid cooled dissipator is shown in Figure 214.

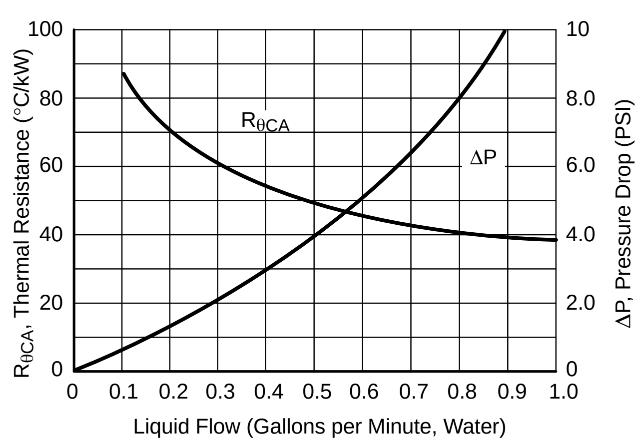


Figure 214. Typical Performance of a Liquid Cooled Mounting Surface

Component Placement

Traditionally, placement techniques have focused on improving routability based on minimizing the total wirelength between interconnected components. However, electronic card assembly (ECA) reliability, which can be measured in terms of time to failure, cycles to failure, or the failure rates of the individual components, the interconnections, and the printed wiring board (PWB), is also affected by component placement.

The thermal characteristics of an electronic system are dependent on both the thermal management techniques being employed and the positioning of the heat dissipating components. When considering placement, an examination of both the cooling methods and the failure mechanisms is required.

In the case where the model for the total failure rate, $h_T(t)$, is time independent, the positioning of the components on the board requires minimizing, as follows:

$$h_T = \sum_{i=1}^{n} h_i(T_i, \Delta T_i) \qquad (13.13)$$

where h_i is the failure rate of the i–th part.

On the other hand, the method used in examining failures due to time dependent failure is based on minimizing the time, or the number of cycles to failure, N_f For example, if the failure mechanism is fatigue caused by temperature cycling, then the number of cycles to failure is often modeled by the Manson–Coffin formulation and has the form:

$$N_f = A(K\Delta T)^n \qquad (13.14)$$

where A, K, and n are based on environmental, geometric, and material properties. The constant n represents the slope on a log–log plot of N_i versus $(K\Delta T)$ and has a negative value.

The optimal reliability method used in failure rate placement cannot always be employed because maximizing the sum of the individual number of cycles to failure does not necessarily ensure the best situation for the component with the lowest cycle to failure component is maximized. Therefore, placement of components based on fatigue failures requires that the components with minimum cycles to failure are successively maximized. Mathematically stated the placement problem becomes

$$\mathrm{Max}(\mathrm{Min}N_{fi}) \text{ for } i = 1, \ldots, n \qquad (13.15)$$

The placement of semiconductors on the heatsink is also important in order to achieve the lowest temperature rise. Figure 215 shows preferred positions. It is generally more economical to use one heatsink with several properly placed transistors than to use individual heatsinks. Cooling efficiency increases and the unit cost decreases under such conditions. Details of placement methods can be found in references [1,2,3].

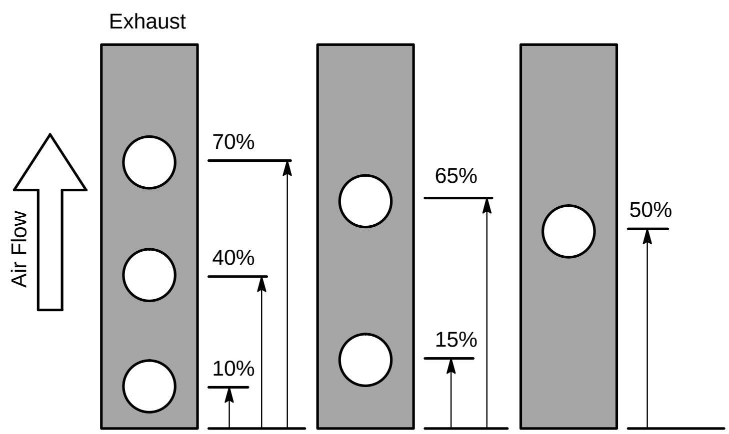


Figure 215. Proper Rectifier Placement for Components having Equal Power Dissipation. Located in this fashion, each semiconductor will operate at the same case temperature and maximize the heatsink performance. In natural convection, mounting must be vertical.

References

1. M. Pecht, *Handbook of Electronic Packaging,* Marcel Dekker, N.Y 1991.
2. D. Steinberg, Cooling Techniques for Electronic Equipment, John Wiley & Sons, N.Y., 1980.
3. Osterman, M. and Pecht, M., “Placement for Reliability and Routability of Convectively Cooled PWBs,” *IEEE Transactions on Computer–aided Design of Integrated Circuits and Systems,* Vol. 9, No. 7, July 1990, pp. 734–744.

Chapter 14

Assembly Considerations for Printed Circuit Boards

Assembly Considerations for Printed Circuit Boards

A rectifier diode package style is chosen based upon the power being dissipated and the assembly technique. Generally, assembly costs are lower when the diode package is soldered directly to the printed circuit board (PCB), using either insertion mount or surface mount, rather than mounting with screws, clips, or other hardware on a separate heatsink. However, the power dissipation capability of conventional PCBs is limited because of the board's relatively poor thermal conductivity. With conventional glass–epoxy boards, component power dissipation is typically limited to 3 W, while more costly metal–backed boards allow dissipations up to about 10 W. The metal–backed board can be attached to a heatsink for even higher dissipation, but this is not often done because the temperature of all components on the board is raised to approximately that of the case of the power components. Thus, in high–power applications, the rectifier diode is usually mounted directly to a heatsink with wires used to connect the component to the PCB.

Many early–life field failures have been traced to faulty mounting procedures which cracked the die, cracked the package case impairing its moisture resistance, or mounted the package in a way that increased its thermal resistance (reduced its ability to transfer the heat away from the die). The objective of a good mounting procedure is to secure the package without causing any physical damage and to provide a low thermal resistance path to the heatsink material. In the sections to follow, mounting considerations are explained for insertion or through–hole PCBs and surface mount PCBs.

Insertion Mounting

Although insertion or through–hole printed circuit boards have been the industry standard since the 1960s, errors are still made in mounting components, particularly when appreciable power must be dissipated. Thermal capabilities are often misjudged and mechanical damage can occur.

For insertion–mounted components in a free convection environment, the primary path of heat transfer is through the leads as discussed in Chapter 2. Sufficient board metal needs to surround the lead insertion hole to act as a heatsink. A circular pattern is most efficient per unit area because the thin board metal has a fairly high thermal resistance. Sometimes semiconductor manufacturers provide information as shown in Chapter 2 that may prove helpful. Otherwise, a required pad area can be calculated using the principles discussed in Chapter 13. Before committing a board assembly to production, the diode lead temperature should be measured using a fine wire thermocouple under worst–case conditions.

Although the leads of axial–leaded and tab–mount packages are somewhat flexible and can be formed for insertion mounting, it is not a recommended procedure to do so because of the risk of damage to the package or die. Often, a mechanical arrangement can be chosen that makes lead–forming unnecessary. Many lead–forming options are available from the manufacturers upon special request. Preformed leads remove the user's risk of device damage caused by bending.

If, however, lead–forming is done by the user, several basic considerations should be observed, When bending the lead, support must be placed between the point of bending and the package. For forming small quantities of units, a pair of pliers may be used to clamp the leads at the case, while bending with the fingers or another pair of pliers. For production quantities, a suitable fixture should be made that also includes lead cutting, thereby eliminating repeated handling of the device.

The following general rules should be observed to avoid damage to the package:

1. A lead bend radius greater than 1/16 inch is advisable.
2. No twisting of leads should be done at the case.
3. No axial motion of the lead should be allowed with respect to the case.

Leads are not designed to withstand excessive axial pull. Force in this direction greater than 4 lb may result in permanent damage to the device. If the mounting arrangement imposes axial stress on the leads, a condition that may be caused by thermal cycling, some method of strain relief should be devised, such as an expansion elbow formed in the lead. When wires are used for connections, care should be exercised to assure that movement of the wire does not cause movement of the lead at the lead–to–case junctions. Highly flexible or braided wires are good for providing strain relief. Wire–wrapping of the leads is permissible, if the lead is restrained between the plastic case and the point of the wrapping.

The leads may be soldered providing specifications are observed. Typically, the maximum soldering temperature for a plastic encapsulated part must not exceed 240°C, and must be applied for not more than *5* sec at a distance greater than 1/8 inch from the case. These numbers can vary depending upon the package type and manufacturer. In all cases, the manufacturer's data sheet should be consulted for the actual specifications. Glass or metal packaged parts usually allow higher temperatures. A removable heat dissipator, clamped on the lead near the body during soldering, reduces the chance of damage when control of temperature and time is not precise, such as with hand soldering. Rosin or activated rosin fluxes are preferred while organic or acid fluxes should not be used.

Pins, leads, and tabs must be handled and connected properly to avoid undue mechanical stress that could cause semiconductor failure. Change in mechanical dimensions as a result of thermal cycling over operating temperature extremes must be considered.

The pins and lugs of metal packaged devices using glass–to–metal seals are not designed to handle any significant stress. If abused, the seals can crack, destroying the hermetic seal. Wires may be attached using sockets, crimp connectors, or solder, provided the data sheet ratings are observed. When wires are attached directly to the terminals, flexible or braided leads are recommended to provide strain relief.

The screw terminals of modules such as the POWERTAP™ package look deceptively rugged. Since the flange base is mounted to a rigid heatsink, the connection to the terminals must allow some flexibility. A rigid buss bar should not be bolted to terminals. Lugs with braid are preferred.

In all cases, the manufacturer’s data sheets should be consulted for recommended mounting procedures and torque specifications. Many packages have special requirements.

Surface–Mount Technology

Surface–mount technology (SMT) began in the early 1970s when electronic manufacturers began to mount miniaturized components directly on the surface of printed circuit boards, a technique evolving from thick film hybrids. SMT is being used increasingly to meet the electronic industry’s demand for boards that are smaller, cheaper, and more reliable.

In surface mounting, the components are soldered directly on top of solder pads called footprints, rather than inserting the leads into holes. Surface–mount devices (SMDs) can either be leaded or leadless. The footprint conforms to the lead layout of the surface–mount component. Two common surface–mount footprints are shown in Figure 216. A comparison between insertion–mount and surface–mount is shown in Figure 217. Surface mounting has several advantages over the insertion–mount method. For example, SMDs are typically 30% to 60% smaller than their insertion–mount versions. With the smaller package size and the elimination of the through–hole, board densities can dramatically increase. The use of surface mount also allows components to be placed on both sides of the PCB. The use of surface–mount chip capacitors, resistors, and semiconductors could theoretically give SMT boards packaging densities similar to hybrids. All this increased density results in lower parasitic inductances and capacitances. Smaller and denser PCBs result in reduced system costs by factoring in effects such as cabinet size, connectors, and cabling.

The current component industry trend is to make all popular leaded products available in surface–mount packaging. This gives the designers the ability to manufacture complete surface–mount assemblies as opposed to hybrid boards–that is, boards which contain both surface–mount and insertion–mount components. Selecting a suitable surface–mounted rectifier diode equivalent to a leaded device generally requires more than a casual look, as the designer may be confronted by a number of different packages.

A number of industry committees play a role in communicating the importance of standardization to the component manufacturer. The JEDEC, ETA, IPC, and SMTA set the standards in the United States while the ElM is the counterpart of JEDEC in Japan. These committees set standards in packaging, outline dimensions, pin spacings, tape and reel specifications, etc.

In conjunction with the industry trend to use automatic placement equipment for surface–mount components, most surface–mount semiconductors are packaged in the industry–standard tape and reel format per EIA 481. The current packaging method is a conductive plastic tape with embossed cavities that serve as a pocket for the individual device. These tape sizes range from 8 to 24 mm depending upon the size of the device. A sealing tape is then applied to retain the device. The reels are typically 7 or 13 inch, the 7 inch reel handling a smaller quantity of devices. Some of the larger packages, such as the D2PAK, may also be packaged in plastic tubes which are sometimes referred to as rails or sticks.

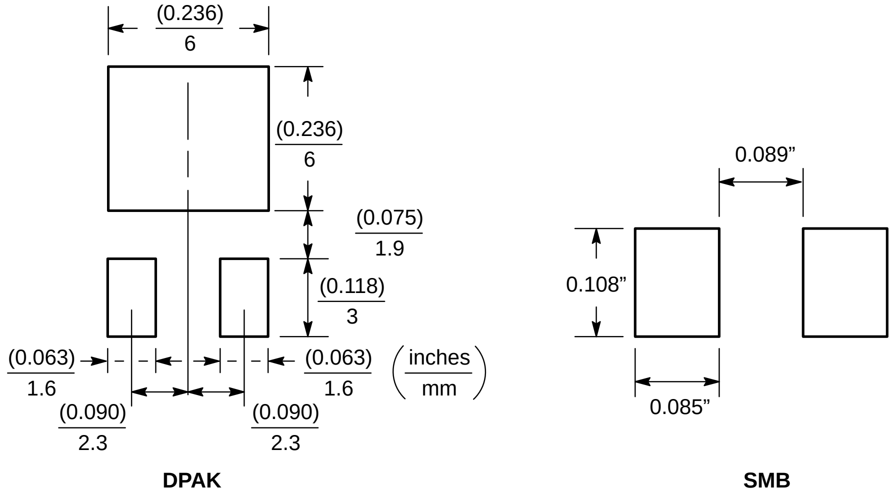


Figure 216. Typical Footprints of Surface–Mount Rectifiers

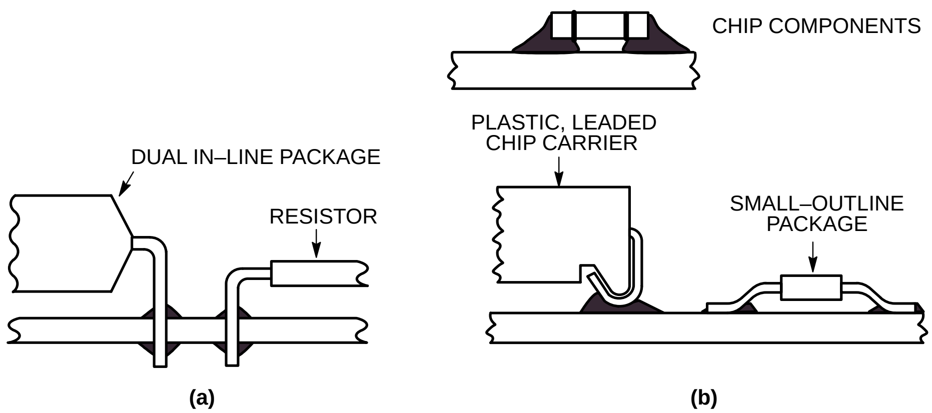


Figure 217. Comparison of (a) Insertion–mount and (b) Surface–mount PCB

Surface–Mount Rectifier Packages

A designer building a surface–mount assembly has a choice of a number of surface–mount packages offered by the semiconductor industry. Depending on the number of leads, these packages can house either single or dual (center–tap) rectifier diodes. There are leadless diodes, SOT–type packages such as SOT–89 and SOT–223, and the SMB and SMC packages which are two leaded packages designed specifically for diodes. For higher power levels, there are the DPAK and the D^2PAK. Some of the more common surface–mount rectifier packages and their general characteristics are shown in Table 30.

Two common types of lead forms are in use on rectifier SMDs. SOTs, DPAKS, and D^2PAKs have “gull–wings”; SMBs and SMCs have the “J” bend. Both of these configurations are shown in Figure 218. Gull–winged leads are more easily probed by test leads, and gull–winged packages are easily handled by automated pick–and–place equipment. Packages with J–bend leads have smaller footprints and take up less real estate on the printed circuit board. The disadvantage is that their solder joints are not easily inspected and test points must be provided to access the leads.

The power dissipation ratings listed in Table 30 assumes that the packages are mounted to a typical glass epoxy board with the minimum recommended size footprints (as shown on the manufacturer’s data sheet) at an ambient temperature of 25°C. Higher power dissipation can be achieved by providing improved heat dissipation through the use of metal–backed boards and ceramic substrates.

Although the term “SOT” stands for small outline transistor, the packages are also used for rectifier products. These are plastic rectangular–shaped packages of varying sizes with gull–wing leads. The number of leads is three for the SOT–23, and four for the SOT–89 and the SOT–223. The power dissipation ratings range from 225 mW for the SOT–23 up to 830 mW for the SOT–223.

The SC–59 package is similar to the SOT–23 and is common in Japan. This package is very close to the SOT–23 in size and has the same thermal and electrical characteristics. Despite similarities, the SC–59 requires a different footprint.

For very high–density, low–power applications, the Japanese have developed several packages smaller than the SOT–23. These packages are the SC–70 and SC–90. The SC–90 is about a third and the SC–70 about half the size of the SOT–23, with power dissipation ratings of 125 and 150 mW respectively. The SC–70 package is also manufactured in five– and six–leaded versions to accommodate diodes in a number of different configurations using monolithic chips. Due to their small size and resulting handling difficulties, these packages are slowly beginning to gain acceptance outside Japan.

SMB and SMC packages resemble “SO” type integrated circuit packages except for wide leads bent in a modified J–bend pattern. The packages have a maximum power dissipation rating of 0.8 and 1.2 W, respectively. Although physically larger, the SMB fits the same footprint as the l–W MELF. The SMB package is an alternative to the leadless MELF and, due to its shape, it is generally easier to handle in manufacturing.

The DPAK developed by ON Semiconductor was the first true power package developed for surface–mount applications. The DPAK resembles a miniature TO–220. In comparison with other plastic power packages, the DPAK is physically smaller than a D^2PAK yet larger than an SMB and SMC. The DPAK allows for a large selection of higher power surface–mount devices since it can house a die size considerably larger than the MELF and SOT type packages. The largest high–power surface mount package presently available is the D2PAK, a package capable of handling the same size die as the TO–220 package with a power dissipation rating of 2.5 W. Both the DPAK and the D^2PAK have gull–winged leadforms with the center lead clipped, leaving the tab for the anode connection.

Prior to the introduction of the DPAK and the D^2PAK, the conventional TO–220 package was used for surface mounting by forming the leads into a gull wing lead form and, sometimes, shearing the tab. (Shearing the tab is a very questionable procedure that often causes a fracture in the die that in turn leads to a latent failure.)

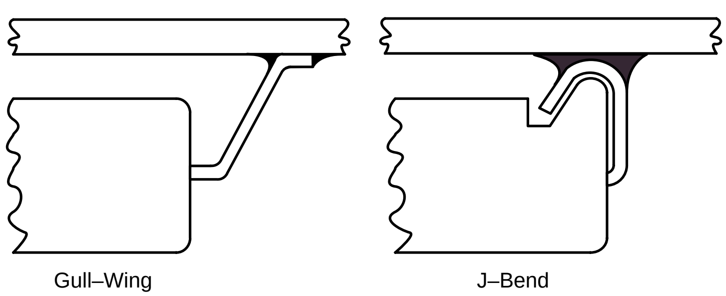
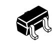
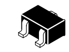
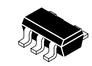
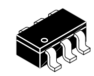
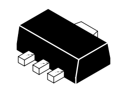
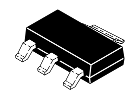
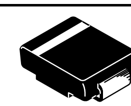
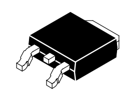
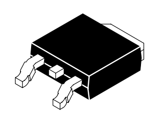


Figure 218. Lead Variations in Surface–mount Components

Table 30. Surface–Mount Rectifier Packages

Package Destination	General Package Description	Max # Leads	Max Die Size (Mils)	Max PD (Watts)	Description (Not To Scale)
SC–90	Low Power, Plastic – Gull–Wing Leads	3	15 x 15	0.125	
SC–70	Low Power, Plastic – Gull–Wing Leads	6	21 x 21	0.15	
SOT–23 (TO–236)	Low Power, Plastic – Gull–Wing Leads	3	25 x 25	0.225	
SC–59	Low Power, Plastic – Gull–Wing Leads	6	25 x 25	0.225	
SOT–89 (TO–243AA)	Low Power, Plastic – Gull–Wing Leads	4	60 x 60	0.6	
SOT–223 (TO–261AA)	Medium Power, Plastic – Gull–Wing Leads	4	90 x 90	0.83	
SMB (DO–214AA)	Medium Power, Plastic – J–Leads	2	60 x 60	0.8	
SMC (DO–214AB)	High Power, Plastic – J–Leads	2	104 x 104	1.2	
DPAK (TO–252)	High Power, Plastic – Gull–Wing Leads and Tab	3	107 x 107	1.75	
D^2PAK (TO–263)	High Power, Plastic – Gull–Wing Leads and Tab	3	170 x 220	2.5	

Thermal Management of Surface–Mount Rectifiers

One of the major considerations when doing a new design is removing the heat from an assembly. The problem becomes even more severe when attempting miniaturization through the use of surface–mount technology. The problem is magnified to a greater extent when placing power surface–mount components such as the DPAK or a D^2PAK on a surface–mount board assembly. Thermal characteristics of surface–mount packages are a major consideration since an increase in junction temperature T_J can have an adverse effect on the long term operating life of the component.

The surface–mount rectifier is soldered to a footprint as shown on the data sheet. This footprint is recommended to achieve the power dissipation ratings as shown on the data sheet. By using a larger footprint and more thermally conductive PCB or substrate materials, the power dissipation capability can be increased. The dedicated amount of PCB area should be based upon thermal management of the circuit. If it is too large, it will defeat the whole concept of using surface mount; if it is too small, it will limit the power–handling capability of the device.

The designer has the choice of either allocating a larger area of copper clad to the footprint or using the smaller and thermally more efficient but more costly ceramics and metal–backed board materials. Designing more complex mechanical thermal assemblies can raise the power–handling capabilities of these packages. The use of higher power components will more than likely dictate the use of such alternate board materials in order to properly manage the thermal characteristics of the assembly.

When using surface–mount rectifiers, the board material may be a glass epoxy, a ceramic substrate, or a metal–backed board to which the device is soldered. The sink–to–ambient thermal impedance will be a function of the substrate material, the size of the footprint available for heat spreading, the proximity of other additional thermal loads on the board, and the velocity of air flow (in the case of forced air cooling). The heat–sinking capacity of the board or substrate will depend on the thermal conductivity of the board material. The reduction in thermal resistance

obtainable will depend on board material, thickness, and air velocity across the board. The use of alternate (non–FR–4) circuit board materials opens up a whole range of capabilities for using the DPAK and D2PAK packages at higher power levels in surface–mount applications. Thorough testing of the prototype is the only way to prove the adequacy of a particular thermal design.

Manufacturers of PCB and substrate materials are constantly working to improve the thermal efficiency of their materials and to develop new materials.

Board Materials

The printed circuit board material for surface–mount components plays a crucial role in ensuring the electrical, thermal, and mechanical reliability of the electronic assembly. There are many types of available materials and it is important to carefully evaluate the important board properties in light of the final application of the product, before a board material can be selected in a cost–effective manner. Each material has a set of properties with particular advantages and disadvantages, as seen in Table 31. No material will satisfy all design needs or applications and compromises in material properties should be sought that offer the best tailoring to suit end–product cost–effectiveness.

Printed circuit boards vary from very basic single–sided boards to very sophisticated multilayer constraining–core structures. In high–reliability applications, thermal management and solder joint reliability concerns often preclude the use of conventional glass epoxy PCB materials. However, there are some selection criteria that are common to all PCB materials as shown in Table 32 which lists design parameters and material properties that affect system performance. Also, Table 33 lists the properties of the most common board materials.

When components are surface mounted, solder acts as both the electrical and the mechanical attachment medium. Mounted to traditional printed circuit board materials, such as glass epoxy, the thermal expansion mismatch between the component and the board can cause solder joint stress failures when a wide operational temperature range is encountered. This problem becomes more apparent with larger component package sizes as the thermal expansion mismatch stress is proportional to package size. Equations are available (see for example IPC–SM–785, Guidelines for Accelerated Reliability Testing of Surface Mounted Solder Attachments) for predicting the life of the solder attaches but for the case of normal rectifier packages, the package dimensions are small enough that thermal expansion mismatch fatigue is not a dominant failure mechanism.

With higher packaging densities and larger individual dies, surface–mount PCBs are often required to remove more heat per unit area than with conventional insertion–mount PCBs. As discussed in Chapter 2, the primary path of heat transfer for SMDs is through the leads, with the exception of the DPAK and the D2PAK. Unfortunately, conventional organic base board materials are relative poor thermal conductors and can require heatsinks or sophisticated cooling mechanisms for thermal management, with resultant penalties in system weight and size. Again, before committing a board assembly to production, the diode lead temperature should be measured using a fine wire thermocouple, under worst–case conditions.

Table 33 summarizes some of the more important properties of board materials. However, for surface–mounting purposes, the properties that most significantly influence the selection of board material, besides thermal conductivity, are glass transition temperature (Tg) and coefficient of thermal expansion (CTE). The simplest definition of glass transition temperature is just what its name suggests: the temperature at which the material changes from a “glass–like” structure to a soft rubber–like consistency. At this temperature, significant changes occur in the mechanical properties of the material. These changes can cause process, handling, and performance problems.

The coefficient of thermal expansion is a measure of how much a material expands per degree of temperature change and is expressed in units of inch/inch per/°C or parts per million/°C. The reinforcing woven fabric materials used in typical multilayer board laminates have a lower CTE than the resin. As a result, the composite laminates have their expansion restrained in the in–plane directions and, due to this restraint, the expansion of the laminate in the out–of–plane direction can be increased over that of the resin alone. Above the Tg, this out–of–plane expansion CTE can increase as much as fivefold and result in substantial strains in plated–through holes.

In general, PCBs will fit into one of four basic categories of construction: organic base material, nonorganic base material, constraining core, and supporting plane or metal–backed boards.

The most common board material in use is FR–4, an epoxy fiberglass laminate. Alternative resins to epoxy are the polyimide, multifunctional epoxies (FR–5). Driven primarily by the military requirements for improved reliability, in–service thermal performance, and field repairability, the polyimides have evolved as the most commonly used material in the high–performance laminate market. The reason the polyimide resins are used is not that they have a better CTE than epoxies, but that they have a substantially higher Tg. Most solder reflow operations occur below the polyimide Tg, not incurring the problems of the high out–of–plane expansion and resin softening as with epoxies. Thus, the problems associated with “pad lifting” during processing is minimized, and the board repairability is enhanced. BT–type resins can be used in place of polyimide at a lower cost.

New fabric–reinforcing materials such as quartz and aramid fiber fabrics are being used to improve the inplane CTE characteristics of boards over those obtained with fiber glass. Unfortunately, due to the extra cost of the materials as

well as the extra processing costs, such as in drill–bit wear, these newer materials have not yet reached general use.

Nonorganic or ceramic–based boards have come directly from hybrid circuit technology with its associated capability for extremely high interconnection densities. These boards have a lower CTE, no worries of a glass transition temperature, and better thermal conductivity but cost more than typical organic–based boards.

Constraining core materials such as copper–Invar–copper or copper–graphite have been developed for use in multilayer PCBs primarily to tailor to the boards overall thermal properties. By varying the ratio of high thermal expansion copper to low thermal expansion Invar, the copper–clad Invar core can be adjusted from 3 to 10 ppm/°C. Figure 219 shows a typical multilayer board with copper–clad Invar power and ground planes acting as constraining cores. The overall CTE of the structure can be tailored by varying the composition and thickness of the constraining cores.

Often metal cores or backings are used as conduction plates in a PCB only to improve the overall thermal characteristics of the board. Figure 220 is a schematic of a simple one–layer metal–backed board specifically designed for improved thermal performance for power SMDs. Using an aluminum base plate with a thermally conductive dielectric layer, which is then overlaid by the copper circuit layer, the component power dissipations can be easily doubled. More complex multilayer laminates with “blind” or “buried” vias can also be bonded to conduction plates.

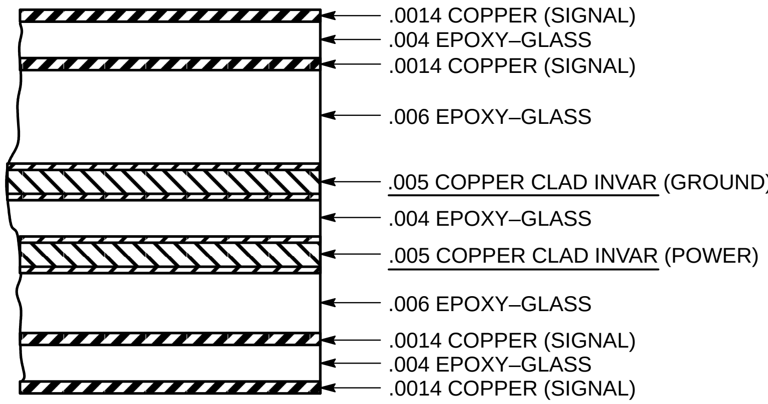
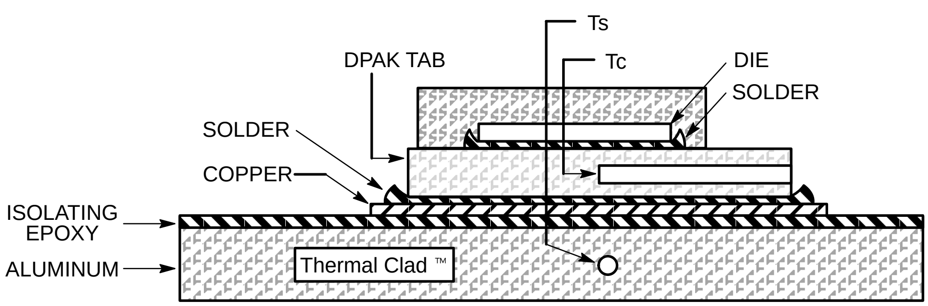


Figure 219. Typical Multilayer Board with Copper–Clad Invar Power and Ground Planes Acting as Constraining Cores



Thermal Clad™ is a trademark of the Bergquist Company

Figure 220. Schematic of a Simple One–Layer Metal–Backed Board Specifically Designed for Improved Thermal Performance for Power SMDs

Table 31. Characteristics of Board Materials

Type	Major Advantages	Major Disadvantages	Comments
Organic base substrate Epoxy Fiberglass	Substrate size, weight, reworkable, dielectric properties, conventional board processing	Thermal conductivity, X, Y, and Z axis CTE	Because of its high X–Y plane, CTE, should be limited to environments and applications with small changes in temperature and/or small packages
Polyimide Fiberglass	Same as epoxy fiberglass plus high–temperature Z axis CTE, substrate size, weight, reworkable, dielectric properties	Thermal conductivity, X and Y axis CTE, moisture absorption	Same as epoxy fiberglass
Epoxy Aramid Fiber	Same as epoxy fiberglass, X–Y axis CTE, substrate size, lightest weight, reworkable, dielectric properties	Thermal conductivity, X and Y axis CTE, resin microcracking, Z axis CTE, water absorption	Volume fraction of fiber can be controlled to tailor X–Y CTE; resin selection critical to reducing resin microcracks
Polyimide Aramid Fiber	Same as epoxy aramid fiber, Z axis CTE, substrate size, weight, reworkable, dielectric properties	Thermal conductivity, X and Y axis CTE, resin microcracking, water absorption	Same as epoxy aramid fiber
Polyimide Quartz (fused silica)	Same as polyimide aramid fiber. Z axis CTE, substrate size, weight, reworkable, dielectric properties	Thermal conductivity, X and Y axis CTE, Z axis CTE, drilling, availability, cost, low resin content required	Volume fraction of fiber can be controlled to tailor X–Y CTE; drill wearout higher than with fiberglass
Fiberglass/Aramid Composite Fiber	Same as polyimide aramid fiber, no surface microcracks, Z axis CTE, substrate size, weight, reworkable, dielectric properties	Thermal conductivity, X and Y axis CTE, water absorption, process solution entrapment	Resin microcracks confined to internal layers and cannot damage external circuitry
Fiberglass/Teflon Laminates	Dielectric constant, high temperature	Same as epoxy fiberglass, low–temperature stability, thermal conductivity, X and Y axis CTE	Suitable for high–speed logic applications, same as epoxy fiberglass
Flexible Dielectric	Light weight, minimal concern to CTE, configuration flexibility	Size	Rigid–flexible boards offer trade–off compromises
Thermoplastic	3–D configurations, low high–volume cost	High injection–molding setup costs	Relatively new for these applications
Nonorganic Base Alumina (ceramic)	CTE, thermal conductivity, conventional thick–film or thin–film processing, integrated resistors	Substrate size, rework limitations, weight, cost, brittle, dielectric constant	Most widely used for hybrid circuit technology
Supporting Plane Printed board bonded to plane support (metal or nonmetal)	Substrate size, reworkability, dielectric properties, conventional board processing, X–Y. axis CTE, stiffness, shielding, cooling	Weight	The thickness/CTE of the metal core can be varied along with the board thickness, to tailor the overall CTE of the composite
Sequential processed board with supporting plane core	Same as board bonded to supporting plane	Weight	Same as board bonded to supporting plane
Discrete Wire	High–speed interconnections; good thermal and electrical features	Licensed process, requires special equipment	Same as board bonded to low–expansion metal support plane
Constraining Core Porcelainized copper–clad invar	Same as alumina	Reworkability, compatible thick–film materials	Thick–film materials still under development
Printed board bonded with contraining metal core	Same as board bonded to supporting plane	Weight, internal layer registration	Same as board bonded to supporting plane
Printed board bonded to low–expansion graphite fiber core	Same as board bonded to low–expansion metal cores, stiffness, thermal conductivity, low weight	Cost	The thickness of the graphite and board can be varied to tailor the overall CTE of the composite
Compliant layer structures	Substrate size, dielectric properties, X–Y axis CTE	Z axis CTE, thermal conductivity	Compliant layer absorbs difference in CTE between ceramic package and substrate

Table 32. Board Structure Selection Criteria

Design Parameters	Material Properties								
	Transition Temperature	Coefficient of Thermal Expansion	Thermal Conductivity	Tensile Modulus	Flexural Modules	Dielectric Constant	Volume Resistivity	Surface Resistivity	Moisture Absorption
Temperature and Power Cycling	X	X	X	X					
Vibration				X	X				
Mechanical Shock				X	X				
Temperature and Humidity	X	X				X	X	X	X
Power Density	X		X						
Chip Carrier Size		X		X					
Circuit Density						X	X	X	
Circuit Speed						X	X	X	

Table 33. PCB Board Properties

Material	Material Properties							
	Glass Transition Temperature (°C)	X–Y Coefficient of Thermal Expansion (ppm/°C) (Note 5)	Thermal Conductivity (W/M°C)	X–Y Tensile Modulus (PSI $\times$ 10^6)	Dielectric Constant (at 1.0 MHz)	Volume Resistivity (Ω/cm)	Surface Resistivity (Ω)	Moisture Absorption (%)
Epoxy Fiberglass	125	13–18	0.16	2.5	4.8	10^{12}	10^{13}	0.10
Polyimide Fiberglass	250	12–16	0.35	2.8	4.8	10^{14}	10^{13}	0.35
Epoxy Aramid Fiber	125	6–8	0.12	4.4	3.9	10^{16}	10^{16}	0.85
Polyimide Aramid Fiber	250	3–7	0.15	4.0	3.6	10^{12}	10^{12}	1.50
Polyimide Quartz	250	6–8	0.30	–	4.0	10^{9}	10^{8}	0.50
Fiberglass/Teflon	75	20	0.26	0.2	2.3	10^{10}	10^{11}	1.10
Thermoplastic Resin	190	25–30	–	3–4	10^{17}	10^{13}	NA	–
Alumina–Beryllia	NA	5–7	44.0	8.0	10^{14}	–	–	–
Aluminum (6061 T–6)	NA	23.6	200	10	NA	10^{6}	–	NA
Copper (CDA101)	NA	17.3	400	17	NA	10^{6}	–	–
Copper–clad Invar	NA	3–6	150XY/20Z	17–22	NA	10^{6}	–	NA
Graphite (P–100)	160	–1.15	110XY/1.1Z	37	NA	–	–	–
Graphite (P–75)	160	–0.97	40XY/.7Z	25	NA	–	–	–

2. These materials can be tailored to provide a wide variety of material properties based on resins, core materials, core thickness, and processing methods.
3. The X and Y expansion is controlled by the core material and only the Z axis is free to expand unrestrained. Where the T_g will be the same as the reinforced resin system used.
4. When used, a compliant layer will conform to the CTE of the base material and to the ceramic component, therefore reducing the strain between the component and the P&I structure.
5. Figures are below glass transition temperature, and are dependent on method of measurement and percentage of resin content.
   NA – Not applicable.

Surface–Mount Assembly

Most SMDs have leads which are either solder plated or solder dipped, thus eliminating the need to pretin the leads prior to mounting. Pretinned PCBs are provided by most manufacturers. This aids in the attachment of the SMDs since both the mechanical and electrical connections are made at the footprint by reflowing solder and joining the pans. Unlike the leaded component joints where the leads provided additional mechanical strength, the SMD is dependent on the solder joint alone. This is why good solder joints are so critical to SMT technology.

A good solder joint must meet three basic criteria: good wetting of surfaces, fillet complete, and not too much solder. To meet these requirements the industry adopted mass soldering techniques. The quality of the end product is determined by the measures taken during the design and manufacturing stages. A consistent standard of quality and reliability in the finished product is not attainable with manual methods.

The assembly process involves the placement of the components on the circuit board with a temporary means of fastening to hold them in place. In insertion mounting, the component leads are inserted into the holes and then clinched, automatically aligning the package and holding it in place until soldering. With surface mounting, the packages are more difficult to handle in assembly because they cannot be aligned during the pick–and–place operation. SMDs must be physically held in place either by an adhesive or with solder paste, depending upon the final configuration and soldering process. Adhesives are used to hold in place SMDs that will be undergoing a wave–soldering process. In a reflow process, the solder paste holds the component in place and alignment is not as much of a concern, as the components generally will be pulled into alignment by the surface tension of the molten solder. The first major step in PCB preparation is to “print” the solder paste onto the pads. This is done by a screen printer. This paste provides the necessary solder to form the joint fillets that are so important to electrical and mechanical connections. The component is then placed on the fresh solder paste and the operation is completed through one of several different types of soldering processes. This process melts the solder and bonds the SMD to the PCB. After the soldering operation, the board is cleaned with a solvent and it is then ready for testing.

There are two general classifications of soldering processes: reflow and wave soldering. Wave soldering is the traditional method used for insertion mount components, and it can also be used for SMDs. In wave soldering, the PCB is inverted so the components are suspended under the board and then passed through a molten stationary wave of solder. The solder wave simultaneously supplies the heat and the solder to make the connection between the component and PCB footprint. In reflow soldering, solder paste is applied to the PCB footprint before the component is placed. The assembly is then heated until the solder paste reflows, making the joint. Many reflow soldering techniques have been developed, which vary only in the method of heat transfer used.

With wave soldering, the component orientation with respect to the direction is very important. The surface tension of the solder can prevent molten solder from reaching the downstream side of the component. The height of the component body is also very critical. Components with heights up to 100 mils can typically be wave soldered. However power SMDs such as the D^2PAK with a body height of 175 mils are not suitable for wave soldering because of “shadowing.” Shadowing occurs when the body height shadows the solder pads from the wave and the surface tension of the molten solder prevents the solder from reaching the far end of the component, as shown in Figure 221.

With the increased package density and the reduction of space between conductors, nonwetting, solder bridging, and shadowing are all problems for conventional wave–soldering equipment. The dual wave method was developed to overcome the limitations of conventional wave soldering. The first wave in a dual wave system is a narrow, turbulent jet–like fountain. The function of this wave is to ensure that solder is forced into the component area and that the solder reaches all pads. The narrow length permits gas evolved from either the flux or component attachment adhesives to escape. The turbulent jet flow of the wave provides sufficient pressure to overcome the repelling and hydrodynamic turbulence caused by the component bodies, assuring that all pads are wetted with solder. The second, laminar, wave embodies the characteristics of a conventional wave–soldering machine and completes the formation of solder fillets and removes bridges. However, with increased circuit density and the resulting reduction of space between conductors, reflow soldering techniques are preferred for surface–mount technology.

Vapor–phase reflow is the most common process used for soldering SMDs. The soldering is accomplished by immersing the assembly in boiling vapors. The vapors condense and transfer heat to the board using the latent heat of condensation principle. Since the liquid boils at a temperature higher than the melting point of the solder, the solder reflows. Vapor–phase reflow soldering offers the advantages of rapid heating with precise temperature control which protects heat–sensitive components from thermal damage. Vapor–phase reflow soldering produces even heating regardless of the overall geometry and is an inherently clean operation as only continuously distilled vapors contact the assembly.

A key ingredient in reflow soldering is the solder paste. Solder paste consists of flux impregnated with small nodules of solder. Solvents are added to the paste to provide optimum flow characteristics during screen printing of the paste onto the substrate assembly. Solder paste reflow can be achieved by several techniques including conduction or convection

ovens, resistance soldering, infrared, laser, and vapor phase reflow.

In thermal conduction belt reflow soldering, the heat required for the reflow is conducted through the assembly. This is accomplished by a series of heated platens located under a continuously moving belt. The platens are individually controlled so that the reflow thermal profile can be varied for assemblies with different thermal conductivities and masses. Generally there are at least two stages: a preheat zone to reduce the thermal shock and to drive off the flux solvents, and a reflow zone for the actual soldering. Because the conduction of the heat through the assembly depends upon the contact of the assembly to the belt, a weighted fixture is usually used to apply pressure to multiple points on the assembly. The method is generally suitable for small substrates with SMDs on the upper side only. If a reducing atmosphere, such as forming gas, is used, then solder post cleaning may not be required.

In thermal–convection or hot–air soldering, the assembly is placed on a conveyor that passes it through various heating zones where hot air or gases are blown across the board. Special attention must be paid to the air velocity. Too much air can blow out or upset the solder in addition to displacing components on the board. If there is too little air, poor convective heat transfer will occur.

In resistance soldering, a heated element in physical contact with the joint is used to reflow the solder. Instead of solder paste, an extra thick layer of solder is usually plated directly onto the footprint.

Infrared reflow uses radiation to transfer heat to the solder paste to cause reflow. With a conveyorized infrared oven, the assembly is passed through preheat zones before reaching the high intensity reflow zone. The main difficulty with IR oven heating is obtaining and maintaining a uniform temperature distribution across the assembly due to shadowing and variabilities in surface emissivity and energy absorption. With careful process control, IR reflow soldering can produce excellent results.

Laser soldering relies upon the absorption of laser radiation by a conductive pad to melt and reflow the solder paste. Laser soldering is not a generally used process. It has its place as a specialized process that takes advantage of the fact that, with its localized heating, it performs its function on precisely the area to be processed, and on that area alone.

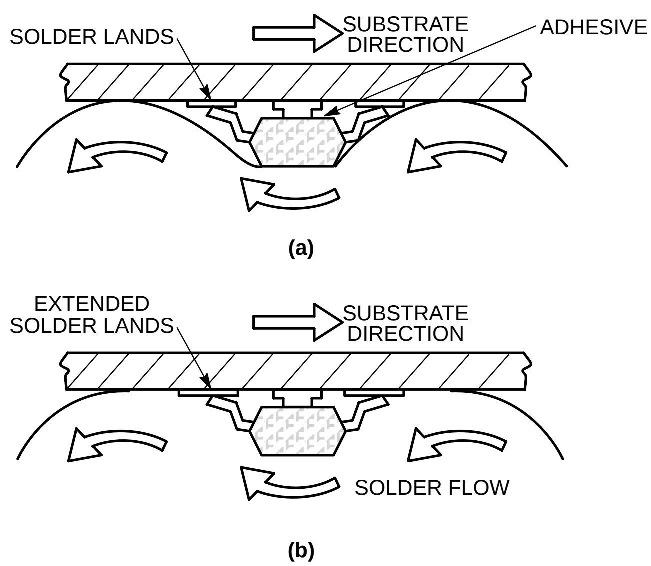


Figure 221. Shadowing in Wave Soldering: (a) Surface Tension and Body Height Can Prevent Molten Solder Reaching the Downstream End of the SMD, the “Shadow Effect”; (b) Extending the Solder Land May Overcome Shadowing

Flux and Post Solder Cleaning

Many factors influence the quality and reliability of the finished product: the soldering process itself, condition of components and substrates being soldered, the choice and application of the flux, and method of post–solder cleaning. Flux selection and post–solder cleaning are becoming the most important yet least understood factors in surface–mount manufacturing today.

The choice and application of the flux, plus the method of flux residue removal from the electronics assemblies are very important in surface–mount technology reliability. The flux removes surface oxides, prevents reoxidation, and assists in the removal of corrosive residues on the substrate during subsequent post–solder cleaning. The flux improves the wettability of the solder joint surfaces. One must be careful, however, not to use a highly activated flux in order to promote rapid wetting to compensate for ill–prepared surfaces. The flux residues are corrosive and a mildly activated flux should be used wherever possible.

The major types of fluxes are shown in Table 34 with the types of cleaning processes suitable for removal of each flux. In the electronics industry rosin/resin fluxes are the most popular.

Rosin is a solid resin obtained from pine trees which, in a pure form but usually with additives, is frequently used as a flux. The main characteristics of a flux are the ability to promote wetting to surfaces, called flux activity (cleaning of surface as solder is reflowing), and the corrosivity of flux residues after soldering. As a general rule the more active the flux, the more corrosive are its residues. In electronic assemblies that are exposed to elevated temperatures and humidity, flux residues from fluxes can corrode the metallurgical tracks on the electronic assemblies and component leads. Post–solder cleaning removes any contamination, such as surface deposits, inclusions, occlusions, or absorbed matter, which may degrade the chemical, physical, or electrical properties of the electronic assemblies.

The removal of fluxes and flux residues after soldering is essential for reliable performance of the electronic assemblies. CFC–113 solvents mixed with alcohols were the solvents of choice for electronic assemblies cleaning prior to concerns about depletion of the ozone layer.

The dilemma brought about by government regulations and the confusion over the multitude of options for specific applications, with no universal replacement for CFC–113, makes it difficult for the designer to select surface–mount technology since the reliability of this technology is very dependent on high–quality cleaning of the electronic assemblies.

Table 34. Types of Fluxes Used for Soldering

Flux	Military Approved**	Organic Solvent	Water	Water with Saponifier	Terpene with water
R (nonactivated rosin)	•a	•		•	•
RMA (rosin, mildly–activated)	•a	•		•	•
RA (rosin, activated)	•a, b	•		•	•
RSA (rosin, super–activated)		•		•	•
WS (water soluble)	c		•	•	
SA (synthetic resin, activated)	d	•			•
Low Solids					

*1 to 10 percent solids.
**U.S. Military
[a]Per the requirements of MIL–F–14256
[b]Military approved for certain applications MIL–F–14258 Qualified Products List; also approved by U.K. Defense Standards.
[c]Approved for use by U.K. Defense Standards.
[d]Conditionally approved for use by U.K. Defense Standards.
Source: Adapted from Markenstein, Howard. 1983. "Solder Flux Developments Expand Choices," *Electronic Packaging and Production,* April 1983, pp. 39–42.

Semiaqueous Cleaning Process

A semiaqueous system is a process in which the contaminants are dissolved from the electronic assemblies and the solvent/residue mixture is rinsed from the surface in a water stage. A solvent–emulsion cleaning system is shown in Figure 222. A solvent stage is the first step in the process, followed by an emulsion stage established in the first rinse stage by drag–out from the solvent stage with a small amount of water added. Overflow from the emulsion stage can be allowed to settle in the decanter vessel. This allows the contaminants generated in the soldering process to be separated from the aqueous effluent. The contaminants collect in the hydrocarbon layer of the decanter. The organics in the water are primarily extracted surfactant from the cleaning agent formula. This is an important distinction from aqueous cleaning, where all dirt and saponifiers are flushed down the drain. This allows the use of a closed–loop water recycle system. With the correct filtration, elimination of water discharge can be possible.

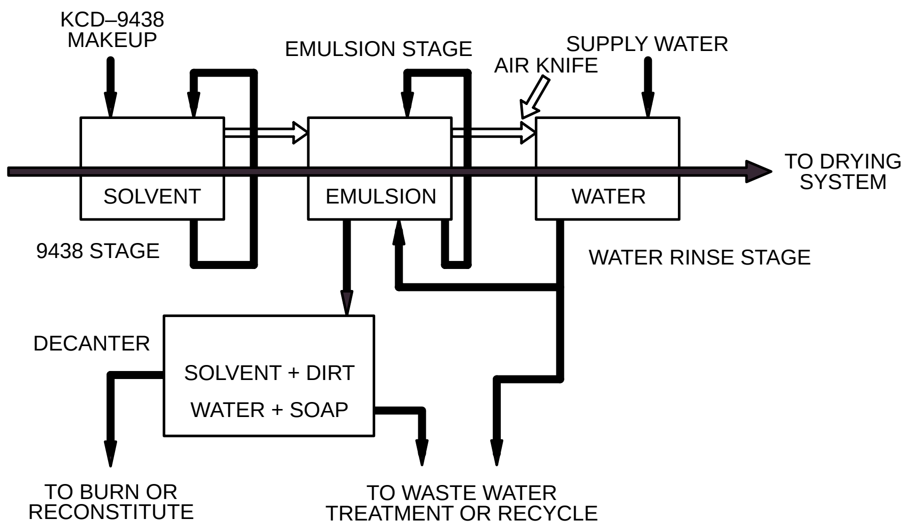


Figure 222. Semiaqueous Cleaning System (courtesy of Du Pont)

Chapter 15

Heatsink Mounting Considerations

Heatsink Mounting Considerations

When the power dissipated in a rectifier diode exceeds approximately 10 W, it is more practical to use an external heatsink rather than a PCB to cool the diode. High–density designs use a cold plate (a plate kept cooled by a flowing fluid) while the finned assemblies discussed in Chapter 13 are used where space allocations are more generous. In any case, proper mounting of the diode to the heatsink is necessary to provide a low resistance path for heat transfer and to secure the package without causing mechanical damage.

In many situations the case of the semiconductor must be electrically isolated from its mounting surface. The isolation material is, to some extent, a thermal isolator as well which raises junction operating temperatures. In addition, the possibility of arc–over problems is introduced if high voltages are present. Various regulating agencies also impose creepage distance specifications which further complicate design. Electrical isolation thus places additional demands upon the materials used and the mounting procedure.

A proper mounting procedure usually necessitates orderly attention to the following:

1. Prepare the mounting surface.
2. Apply a thermal grease (if required).
3. Install the insulator (if electrical isolation is desired).
4. Fasten the assembly.
5. Connect the terminals to the circuit.

Mounting procedures are dependent upon the package. However, packages can be placed into one of several generic classes of packages. As newer packages are developed, it is probable that they will fit into the generic classes defined. Unique requirements are given on data sheets pertaining to the particular package. The following classes are defined:

- Stud mount
- Range mount
- Pressfit
- Plastic–body mount
- Tab mount

Mounting Surface Preparation

Since diodes are cooled by contact between the diode case and the heatsink surface, it is important that the maximum possible area of the diode case is in intimate contact with the heatsink. Consequently, surface flatness and finish should meet minimum standards and the mounting method should ideally distribute pressure evenly over the mating surfaces. Mounting holes and surface treatment must also be considered.

Surface Flatness

In general, the heatsink mounting surface should have a flatness and finish comparable to that of the semiconductor package. In lower power applications, the heatsink surface is satisfactory if it appears flat against a straight edge and is free from deep scratches. In high–power applications, a more detailed examination of the surface is required.

Surface flatness is determined by comparing the variance in height (Δh) of the test specimen to that of a reference standard as indicated in Figure 223. Flatness is normally specified as a fraction of the Total Indicter Reading (TIR). The mounting surface flatness, i.e., Δh/TlR, if less than 4 mils per inch (normal for extruded aluminum) is satisfactory in most cases.

Surface Finish

Surface finish is the average of the deviations both above and below the mean value of surface height. For minimum interface resistance, a finish in the range of 50 to 60 µ–in. is satisfactory; a finer finish is costly to achieve and does not significantly lower contact resistance. Tests using a copper TO–3 package with a typical 32 µ–in. finish, showed that heatsink finishes between 16 and 64 µ–in. caused less than ±2.5% difference in interface thermal resistance when the voids and scratches were filled with a thermal joint compound [1]. Most commercially available cast or extruded heatsinks will require spotfacing when used in high–power applications. In general, milled or machined surfaces are satisfactory if prepared with tools in good working condition.

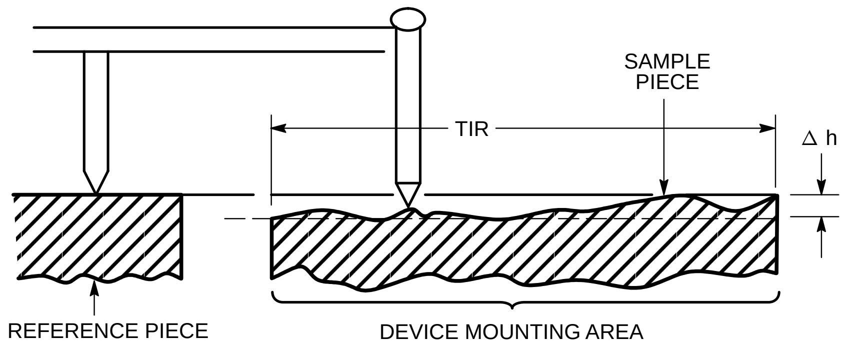


Figure 223. Surface Flatness Measurement

Mounting Holes

Mounting holes generally should only be large enough to allow clearance of the fastener. The larger thick–flange packages having mounting holes removed from the semiconductor die location, such as the TO–3, may successfully be used with larger holes to accommodate an insulating bushing, but many plastic encapsulated packages are intolerant of this condition. For these packages, a smaller screw size must be used so that the hole for the bushing does not exceed the hole in the package.

Punched mounting holes have been a source of trouble because, if not properly done, the area around a punched hole is depressed in the process. This "crater" in the heatsink around the mounting hole can cause two problems. The device can be damaged by distortion of the package as the mounting pressure attempts to conform it to the shape of the heatsink indentation, or the device may only bridge the crater and leave a significant percentage of its heat–dissipating surface out of contact with the heatsink. The first effect may often be detected immediately by visual cracks in the package (if plastic), but usually an unnatural stress is imposed, which results in an early–life failure. The second effect results in hotter operation and may shorten long–term life.

Although punched holes are seldom acceptable in the relatively thick material used for extruded aluminum heatsinks, several manufacturers are capable of properly utilizing the capabilities inherent in both fine–edge blanking or sheared–through holes when applied to sheet metal as commonly used for stamped heatsinks. The holes are pierced using Class A progressive dies mounted on four–post die sets equipped with proper pressure pads and holding fixtures.

When mounting holes are drilled, a general practice with extruded aluminum, surface cleanup is important. Chamfers must be avoided because they reduce heat transfer surface and increase mounting stress. However, the edges must be broken to remove burrs which cause poor contact between device and heatsink and may puncture isolation material.

Surface Treatment

Many aluminum heatsinks are black–anodized to improve radiation ability and prevent corrosion. Anodizing results in significant electrical but negligible thermal insulation. It need only be removed from the mounting area when electrical contact is required. Heatsinks are also available that have a nickel–plated copper insert under the semiconductor mounting area. No treatment of this surface is necessary.

Another treated aluminum finish is indite, or chromate–acid dip which offers low resistance because of its thin surface, yet has good electrical properties because it resists oxidation. It need only be cleaned of the oils and films that collect in the manufacture and storage of the sinks, a practice which should be applied to all heatsinks.

For economy, paint is sometimes used for sinks; removal of the paint where the semiconductor is attached is usually required because of paint's high thermal resistance. However, when it is necessary to insulate the semiconductor package from the heatsink, hard anodized or painted surfaces allow an easy installation for low–voltage applications. Some manufacturers will provide anodized or painted surfaces meeting specific insulation voltage requirements, usually up to 400 V.

It is also necessary that the surface be free from all foreign material, film, and oxide (freshly bared aluminum forms an oxide layer in a few seconds). Immediately prior to assembly, it is a good practice to polish the mounting area with No. 400–600 grit paper or No. 000 steel wool, followed by an acetone or alcohol rinse. If steel wool is used, care must be taken to remove all steel particles to avoid flashover.

Interface Decisions

Even with properly prepared mounting surfaces, when any significant amount of power is being dissipated, something must be done to fill the air voids between mating surfaces in the thermal path. Otherwise, the interface thermal resistance will be unnecessarily high and quite dependent upon the surface finishes and pressure.

For several years, thermal joint compounds, often called grease, have been used in the interface. They have a resistivity of approximately 60°C/W/in. whereas air has 1200°C/W/in. Since surfaces are highly pockmarked with minute voids, use of a compound makes a significant reduction in the interface thermal resistance of the joint. However, the grease causes a number of problems, as discussed in the following section.

To avoid using grease, manufacturers have developed dry conductive and insulating pads to replace the more traditional materials. These pads are conformal and therefore partially fill voids when under pressure.

Thermal Compounds (Grease)

Joint compounds are a formulation of fine zinc or other conductive particles in a silicone oil or other synthetic base fluid, which maintains a grease–like consistency with time and temperature. Since some of these compounds do not spread well, they should be evenly applied in a very thin layer using a spatula or lintless brush, and wiped lightly to remove excess material. Some cyclic rotation of the package will help the compound spread evenly over the entire contact area. Some experimentation is necessary to determine the correct quantity; too little will not fill all the voids, while too much may permit some compound to remain between well–mated metal surfaces where it will substantially increase the thermal resistance of the joint.

To determine the correct amount, several semiconductor samples and heatsinks should be assembled with different amounts of grease applied evenly to one side of each mating surface. When the amount is correct, a very small amount of grease should appear around the perimeter of each mating surface as the assembly is slowly torqued to the recommended value. Examination of a dismantled assembly should reveal even wetting across each mating surface. In production, assemblers should be trained to slowly apply the specified torque even though an excessive amount of grease appears at the edges of mating surfaces. Insufficient torque causes a significant increase in the thermal resistance of the interface.

To prevent accumulation of airborne particulate matter, excess compound should be wiped away using a cloth moistened with acetone or alcohol. These solvents should not contact plastic–encapsulated devices, as they may enter the package and cause a leakage path or carry in substances which might attack the semiconductor chip.

Data showing the effect of compounds on several package types under different mounting conditions is shown in Table 35. The rougher the surface, the more valuable the grease becomes in lowering contact resistance; therefore, when mica insulating washers are used, use of grease is generally mandatory. The joint compound also improves the breakdown rating of the insulator.

The silicone oil used in most greases has been found to evaporate from hot surfaces with time and become deposited on other cooler surfaces. Consequently, manufacturers must determine whether a microscopically–thin coating of silicone oil on the entire assembly will pose any problems. It may be necessary to enclose components using grease. The newer synthetic–base greases show far less tendency to migrate or creep than those made with a silicone oil base. However, their currently observed working temperature range is less, they are slightly poorer on thermal conductivity and dielectric strength, and their cost is higher.

Table 35. Approximate Values for Interface Thermal Resistance Data from Measurements Performed in ON Semiconductor Applications Engineering Laboratory

Dry interface values are subject to a wide variation because of extreme dependence upon surface conditions. Unless otherwise noted, the case temperature is monitored by a thermocouple locate directly under the die reached through a hole in the heatsink.

Package Type and Data		Interface Thermal Resistance (°C/W)						
			Metal–to–Metal		With Insulator			See Note
JEDEC Outlines	Description	Test Torque In–Lb	Dry	Lubed	Dry	Lubed	Type	
DO–203AA, TO–210AA, TO–208AB	10–32 Stud 7/16″	15	0.3	0.2	1.6	0.8	3 mil Mica	
DO–203AB, TO–210AC, TO–208	1/4–28 Stud 11/16″ Hex	25	0.2	0.1	0.8	0.6	5 mil Mica	
DO–208AA	Pressfit, 1/2″	–	0.15	0.1	–	–	–	
TO–204AA (TO–3)	Diamond Flange	6	0.5	0.1	1.3	0.36	3 mil Mica	1
TO–213AA (TO–66)	Diamond Flange	6	1.5	0.5	2.3	0.9	2 mil Mica	
TO–126	Thermopad 1/4″ x 3/8″	6	2.0	1.3	4.3	3.3	2 mil Mica	
TO–220AB	Thermowatt	8	1.2	1.0	3.4	1.6	2 mil Mica	1, 2

1. See Figures 224, 225, and 226 for additional data on TO–3 and TO–220 packages.
2. Screw not insulated. See Figure 234.

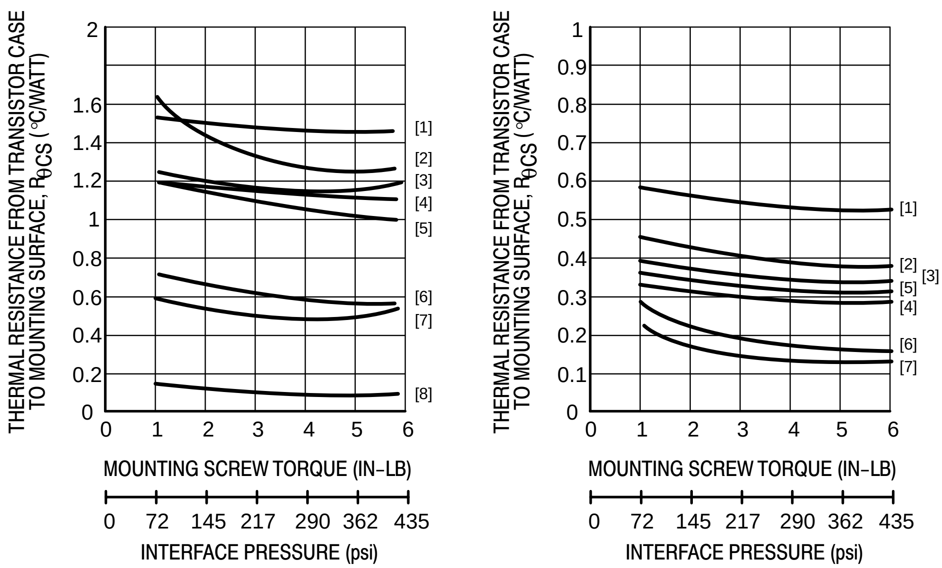
Conductive Pads

Because of the difficulty of assembly using grease and the evaporation problem, some equipment manufacturers will not, or cannot, use grease. To minimize the need for grease, several vendors offer dry conductive pads which approximate performance obtained with grease. Data for a greased bare joint and a joint using Grafoil®, a dry graphite compound, is shown in the data of Figures 224 and 225. Grafoil is claimed to be a replacement for grease when no electrical isolation is required; the data indicates it does indeed perform as well as grease. Another conductive pad available from AAVID is called Kon–Dux™. It is made with a unique, grain–oriented, flake–like structure (patent pending). Highly compressible, it becomes formed to the surface irregularities of both the heatsink and the semiconductor. Manufacturer's data shows it to provide an interface thermal resistance, better than a metal interface with filled silicone grease. Similar dry conductive pads are available from other manufacturers. They are a fairly recent development; long–term problems, if they exist, have not yet become evident.

Insulation Considerations

Since most power semiconductors use vertical device construction, it is common to manufacture power diodes with the anode electrically common to the case; the problem of isolating this terminal from ground is a common one. For lowest overall thermal resistance which is quite important when high power must be dissipated, it is best to isolate the entire heatsink/semiconductor structure from ground, rather than to use an insulator between the semiconductor and the heatsink. Heatsink isolation is not always possible, however, because of EMI requirements, safety reasons, and instances where a chassis serves as a heatsink or where a heatsink is common to several nonisolated packages. In these situations insulators are used to isolate the individual components from the heatsink. Newer packages contain the electrical isolation material within, thereby saving the equipment manufacturer the burden of addressing the isolation problem.

When an insulator is used, thermal grease is of greater importance than with a metal–to–metal contact because two interfaces exist instead of one, and some insulating materials, such as mica, have a hard, markedly uneven surface. With many isolation materials, reduction of interface thermal resistance by 50%–70% is typical when grease is used.



(a) TO–204AA (TO–3)
Without Thermal Grease

(b) TO–204AA (TO–3)
With Thermal Grease

(1) **Thermalfilm**™, 0.002 (0.05) thick.
(2) **Mica**, 0.003 (0.08) thick.
(3) **Mica**, 0.002 (0.05) thick.
(4) **Hard anodized**, 0.020 (0.51) thick.
(5) **Aluminum oxide**, 0.062 (1.57) thick.
(6) **Beryllium oxide**, 0.62 (1.57) thick.
(7) **Bare joint** – no finish.
(8) **Grafoil**®, 0.005 (0.13) thick.*
*__Grafoil__ is not an insulating material.
(Data courtesy of Thermalloy, Inc.)

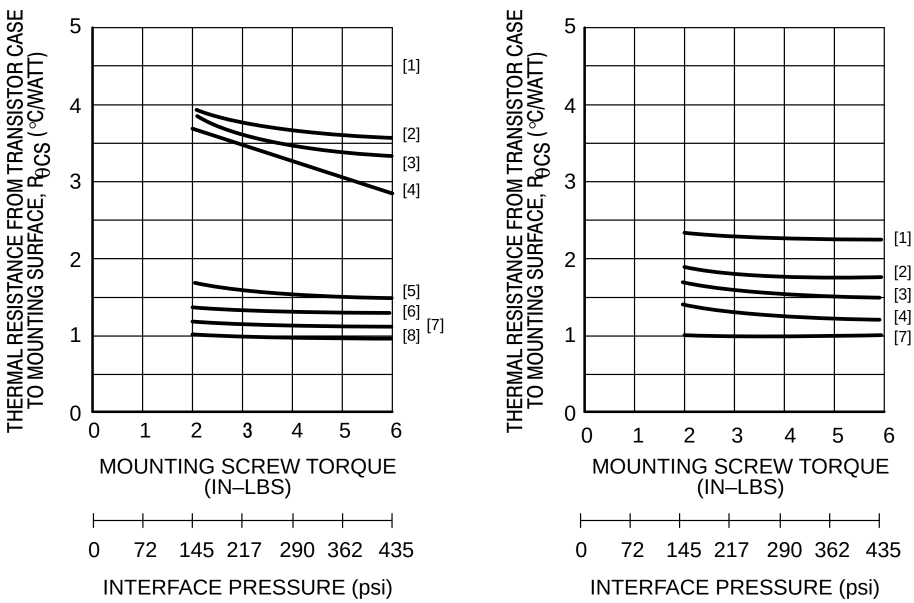
Figure 224. Interface Thermal Resistance for a TO–204 (TO–3) Package

Data showing interface resistance for different insulators and torques applied to TO–3 and TO–220 packages are shown in Figures 224 and 225 for bare and greased surfaces. Similar materials to those shown are available from several manufacturers. It is obvious that with some arrangements, the interface thermal resistance exceeds that of the semiconductor (junction–to–case).

The conclusion from Figures 224 and 225 is that beryllium oxide is unquestionably the best choice. However, it is an expensive choice. (It should not be cut or abraded, as the dust is highly toxic if inhaled.) Thermalfilm™ is a filled polymide material which is used for isolation (variation of Kapton®). It is a popular material for low–power applications because of its low cost, ability to withstand high temperatures, and ease of handling in contrast to mica which chips and flakes easily.

A number of other insulating materials are also shown. They cover a wide range of insulation resistance, thermal resistance, and ease of handling. Mica has been widely used in the past because it offers high breakdown voltage and fairly low thermal resistance at a low cost, but it certainly should be used with grease.

Silicone rubber insulators have gained favor because they are somewhat conformal under pressure. Their ability to fill in most of the metal voids at the interface reduces the need for thermal grease. When first introduced, they suffered from cut–through after a few years in service. The ones presently available have solved this problem by having embedded pads of Kapton® or fiberglass. By comparing Figures 225(a) and 225(b), it can be noted that Thermalsil®, a filled silicone rubber, without grease has about the same interface thermal resistance as greased mica for the TO–220 package.



(a) TO–220
Without Thermal Grease

(b) TO–220
With Thermal Grease

(1) **Thermalfilm**™, 0.022 (0.05) thick.
(2) **Mica**, 0.003, (0.08) thick.
(3) **Mica**, 0.002 (0.05) thick.
(4) **Hard Anodized**, 0.020 (0.51) thick.
(5) **Thermalsil® II**, 0.009 (0.23) thick.
(6) **Thermalsil® III**, 0.006 (0.15) thick.
(7) **Bare Joint** – no finish.
(8) **Grafoil®**, 0.005 (0.13) thick.*
***Grafoil** is not an insulating material.
(Data courtesy of Thermalloy, Inc.)

Figure 225. Interface Thermal Resistance for a TO–220 Package

A number of manufacturers offer silicone rubber insulators. Table 36 shows measured performance of several of these insulators under carefully controlled, nearly identical conditions. The interface thermal resistance extremes are over 2:1 for the various materials. It is also clear that some of the insulators are much more tolerant than others of out–of–flat surfaces. Since the tests were performed, newer products have been introduced. The Bergquist K–10 pad, for example, is described as having about two–thirds the interface resistance of the Sil Pad 1000, which would place its performance close to the Chomerics 1671 pad. AAVID also offers an isolated pad called Rubber–Duc; however, it is only available vulcanized to a heatsink and therefore was not included in the comparison. Published data from AAVID shows $R_{\theta CS}$ below 0.3°C/W for pressures above 500 psi. However, surface flatness and other details are not specified, so a comparison cannot be made with other data in this note.

The thermal resistance of some silicone rubber insulators is sensitive to surface flatness when used under a fairly rigid base package. Data for a TO–3 package insulated with Thermalsil® is shown in Figure 226. Observe that the “worst case” encountered (7.5 mils) yields results having about twice the thermal resistance of the “typical case” (3 mils), for the more thermally conductive insulator. In order for Thermalsil® III to exceed the performance of greased mica, total surface flatness must be under 2 mils, a situation that requires spot finishing.

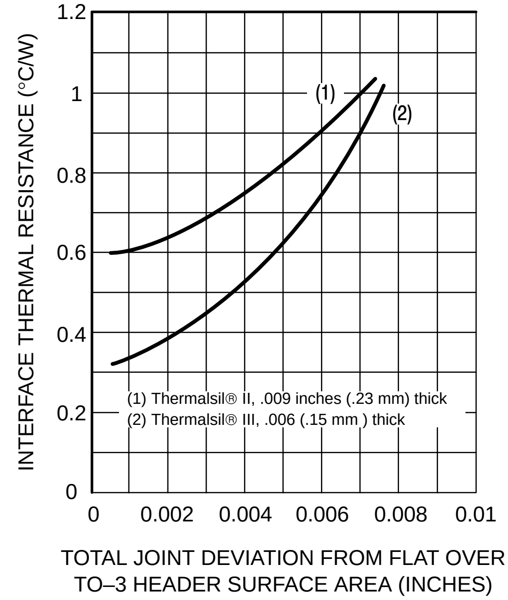


Figure 226. Effect of Total Surface Flatness on Interface Resistance Using Silicon Rubber Insulators

Table 36. Thermal Resistance of Silicone Rubber Pads

Manufacturer	Product	$R_{\theta CS}$ @ 3 Mils*	$R_{\theta CS}$ @ 7.5 Mils*
Wakefield	Delta Pad 173–7	.790	1.175
Berquist	Sir Pad K–4	.752	1.470
Stockwell Rubber	1867	.742	1.015
Berquist	Sil Pad 400–9	.735	1.205
Thermalloy	Thermalsil® II	.680	1.045
Shin–Etsu	TC–30AG	.664	1.260
Berquist	Sil Pad 400–7	.633	1.060
Chomerics	1674	.592	1.190
Wakefield	Delta Pad 174–9	.574	.755
Bergquist	Sil Pad 1000	.529	.935
Ablestik	Thermal Wafers	.500	.990
Thermalloy	Thermalsil® III	.440	1.035
Chomerics	1671	.367	.655

*Test Fixture Deviation from flat. From Thermalloy EIR86–1010.

Silicon rubber insulators have a number of unusual characteristics. Besides being affected by surface flatness and initial contact pressure, time is a factor. For example, in a study of the Cho–Therm® 1688 pad, thermal interface impedance dropped from 0.90°C/W to 0.70°C/W at the end of 1000 hours. Most of the change occurred during the first 200 hours where $R_{\theta CS}$ measured 0.74°C/W. Because the pad "relaxed," the torque on the conventional mounting hardware had decreased to 3 in.–lb from an initial 6 in.–lb. With nonconformal materials, a reduction in torque would have increased the interface thermal resistance.

Because of the difficulties in controlling all variables affecting tests of interface thermal resistance, data from different manufacturers is not in good agreement. Table 37 shows data obtained from two sources. The relative performance is the same, except for mica which varies widely in thickness. At the time of this writing, ASTM Committee D9 is developing a standard for interface measurements.

The conclusion to be drawn from all this data is that some types of silicon rubber pads, mounted dry, will out–perform the commonly used mica with grease. Insulator and assembly cost may be a determining factor in making a selection.

Insulation Resistance

When using insulators, care must be taken to keep the mating surfaces clean. Small particles of foreign matter can puncture the insulation, rendering it useless or seriously lowering its dielectric strength. In addition, particularly when voltages higher than 300 V are encountered, problems with creepage may occur. Dust and other foreign material can shorten creepage distances significantly, so having a clean assembly area is important. Surface roughness and humidity also lower insulation resistance. Use of thermal grease usually raises the breakdown voltage of the insulation system but excess grease must be removed to avoid collecting dust. Because of these factors, which are not amenable to analysis, hi–pot testing should be done on prototypes and a large margin of safety employed.

Insulated Electrode Packages

Because of the nuisance of handling and installing the accessories needed for an insulated semiconductor mounting, equipment manufacturers have longed for cost–effective insulated packages since the 1950s. The first to appear were stud–mount types, which usually have a layer of beryllium oxide between the stud hex and the can. Although effective, the assembly is costly and requires manual mounting and lead wire soldering to terminals on top of the case. Consequently, few rectifiers are available in an insulated stud package. In the late 1980s, a number of electrically isolated parts became available from various semiconductor manufacturers. Insulated parts are generally handled the same as their noninsulated counterparts, but some specify special procedures. With insulated parts it is especially important to consult manufacturers' data sheets.

Table 37. Performance of Silicon Rubber Insulators Tested per MIL–I–49456

Material	Measured Thermal Resistance (°C/W)	
	Thermalloy Data (Note 3)	Berquist Data (Note 4)
Bare Joint, greased	0.033	0.008
BeO, greased	0.082	–
Cho–Therm, 1617	0.233	–
Q Pad (noninsulated)	–	0.009
Sil–Pad, K–10	0.263	0.200
Thermalsil III	0.267	–
Mica, greased	0.329	0.400
Sil–Pad 1000	0.400	0.300
Cho–therm, 1674	0.433	–
Thermalsil II	0.500	–
Sil–Pad 400	0.533	0.440
Sil–Pad K–4	0.583	0.440

3. From Thermalloy EIR 87–1030.
4. From Berquist Data Sheet.

Fastener and Hardware Characteristics

Characteristics of fasteners, associated hardware, and the tools to secure them determine their suitability for use in mounting the various packages. Since many problems have arisen because of improper choices, the basic characteristics of several types of hardware are discussed next.

Compression Hardware

Normal split–ring lock washers are not the best choice for mounting power semiconductors. A typical #6 washer flattens at about 50 lb, whereas 150 to 300 lb is needed for good heat transfer at the interface. A very useful piece of hardware is the conical compression washer, sometimes called a Belleville washer. As shown in Figure 227, it has the ability to maintain a fairly constant pressure over a wide range of its physical deflection–generally 20% to 80%. When installing, the assembler applies torque until the washer depresses to half its original height. (Tests should be run prior to setting up the assembly line to determine the proper torque for the fastener used to achieve 50% deflection.) The washer will absorb any cyclic expansion of the package, insulating washer, or other materials caused by temperature changes. Conical washers are the key to successful mounting of devices requiring strict control of the mounting force or when plastic hardware is used in the mounting scheme. They are used with the large face contacting the packages. A new variation of the conical washer includes it as part of a nut assembly. Called a Sync–Nut™ [2], the patented device can be soldered to a PC board and the semiconductor mounted with a 6/32 machine screw.

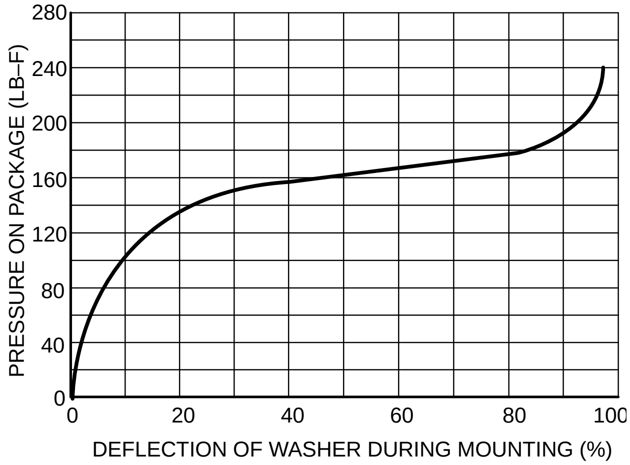


Figure 227. Characteristics of the Conical Compression Washers Designed for Use with Plastic Body–Mounted Semiconductors

Clips

Fast assembly is accomplished with clips. When only a few watts are being dissipated, the small board–mounted or free–standing heat dissipators with an integral clip, offered by several manufacturers, result in a low–cost assembly. When higher power is being handled, a separate clip may be used with larger heatsinks. In order to provide proper pressure, the clip must be specially designed for a particular heatsink thickness and semiconductor package.

Clips are especially popular with plastic packages such as the TO–220. In addition to fast assembly, the clip provides lower interface thermal resistance than other assembly methods when it is designed for proper pressure to bear on the top of the plastic over the die. The TO–220 package usually is lifted up under the die location when mounted with a single fastener through the hole in the tab because of the high pressure at one end.

Machine Screws

Machine screws, conical washers, and nuts (or Sync–Nuts™) can form a trouble–free fastener system for all types of packages which have mounting holes. However, proper torque is necessary. Torque ratings apply when dry; therefore, care must be exercised when using thermal grease to prevent it from getting on the threads as inconsistent torque readings result. Machine screw heads should not directly contact the surface of plastic packages as the screw heads are not sufficiently flat to provide properly distributed force. Without a flat washer under the head, cracking of the plastic case may occur.

Self–Tapping Screws

Under carefully controlled conditions, sheet–metal screws are acceptable. However, during the tapping process with a standard screw, a volcano–like protrusion develops in the metal being threaded; an unacceptable surface that increases the thermal resistance usually results. When standard sheet metal screws are used, they must be used in a clearance hole to engage a speed–nut. If a self–tapping process is desired, the screw type must be used that roll–forms machine screw threads.

Rivets

Rivets are not a recommended fastener for any of the plastic packages. When a rugged metal flange–mount package is being mounted directly to a heatsink, rivets can be used provided press–riveting is used. Crimping force must be applied slowly and evenly. Pop–riveting should never be used because the high crimping force causes deformation of most semiconductor packages. Aluminum rivets are much preferred over steel because less pressure is required to set the rivet and thermal conductivity is improved.

The hollow rivet, or eyelet, is preferred over solid rivets. An adjustable, regulated pressure press is used so that a gradually increasing pressure is used to pan the eyelet. Use of sharp blows could damage the semiconductor die.

Adhesives

Adhesives are available [3] that have coefficients of expansion compatible with copper and aluminum. Highly conductive types are available; a 10–mil layer has approximately 0.3°C/W interface thermal resistance. Different types are offered: high–strength types for nonfield–serviceable systems or low–strength types for field–serviceable systems. Adhesive bonding is attractive when case–mounted parts are used in wave–soldering assembly because thermal greases are not compatible with the conformal coatings used and the greases foul the solder process.

Plastic Hardware

Most plastic materials will flow, but differ widely in this characteristic. When plastic materials form parts of the fastening system, compression washers are highly valuable to assure that the assembly will not loosen with time and temperature cycling. As previously discussed, loss of contact pressure increases interface thermal resistance.

Fastening Techniques

Each of the various classes of packages in use requires different fastening techniques. Details pertaining to each type are discussed in following sections. Some general considerations follow.

To prevent galvanic action from occurring when devices are used on aluminum heatsinks in a corrosive atmosphere, many devices are nickel– or gold–plated. Consequently, precautions must be taken not to mar the finish.

Another factor to be considered is that when a copper–based part is rigidly mounted to an aluminum heatsink, a bimetallic system results that will bend with temperature changes. Not only is the thermal coefficient of expansion different for copper and aluminum, but the temperature gradient through each metal also causes each component to bend. If bending is excessive and the package is rigidly mounted by two or more screws, the semiconductor chip could be damaged. Bending can be minimized by:

1. Mounting the component with the plane between the screws, parallel to the heatsink fins to provide increased stiffness.
2. Allowing the heatsink holes to be a bit oversized so that some slip between surfaces can occur as temperature changes.
3. Using a highly conductive thermal grease or mounting pad between the heatsink and semiconductor to minimize the temperature gradient and allow for movement.

Stud Mount

Parts in the stud–mount classification are shown in Figure 228. Mounting errors with noninsulated stud–mounted parts are generally confined to application of excessive torque or tapping the stud into a threaded heatsink hole. Both these practices may cause a warpage of the hex base which may crack the semiconductor die. The only recommended fastening method is to use a nut and washer; the details are shown in Figure 229.

Insulated electrode packages on a stud–mount base require less hardware. They are mounted the same as their noninsulated counterparts, but care must be exercised to avoid applying a shear or tension stress to the insulation layer, usually a beryllium oxide (BeO) ceramic. This requirement dictates that the leads must be attached to the circuit with flexible wire. In addition, the stud hex should be used to hold the part while the nut is torqued.

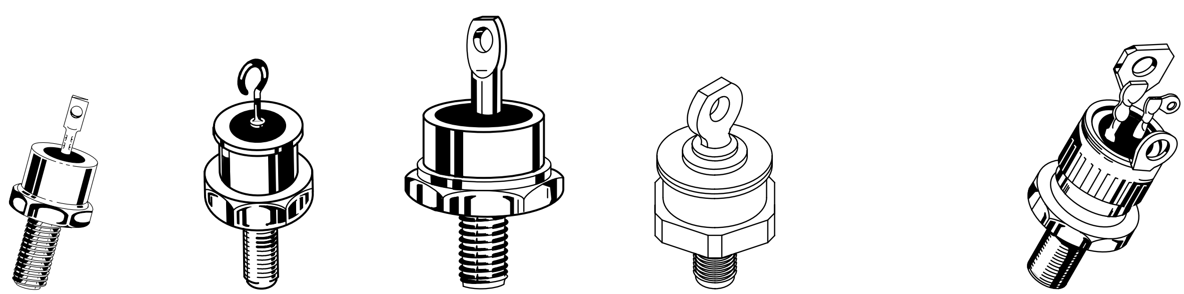
Press Fit

For most applications, the press–fit case should be mounted according to the instructions shown in Figure 230. A special fixture meeting the necessary requirements must be used.

Flange Mount

Rectifier packages which fit into the flange mount category are shown in Figure 231. Few known mounting difficulties exist. The rugged base and distance between die and mounting holes combine to make it extremely difficult to cause any warpage unless mounted on a surface that is badly bowed or unless one side is tightened excessively before the other screw is started. It is, therefore, good practice to alternate tightening of the screws so that pressure is evenly applied. After the screws are finger–tight, the hardware should be torqued to its final specification in at least two sequential steps. A typical mounting installation for a popular flange–type part is shown in Figure 232. Machine screws (preferred), self–tapping screws, eyelets, or rivets may be used to secure the package using guidelines in the previous section, *Fastener and Hardware Characteristics.*

Some packages specify a tightening procedure. For example, with the POWERTAP™ package, Figure 231(b), final torque should be applied first to the center position.



CASE 56–03 DO–203AA (DO–4) | CASE 245 (DO–4) | CASE 257 DO–203AB (DO–5) | CASE 42A (DO–5) | CASE 311

(a) (b)

Figure 228. Stud–Mount Rectifier Packages: (a) Standard nonisolated types; (b) Isolated type

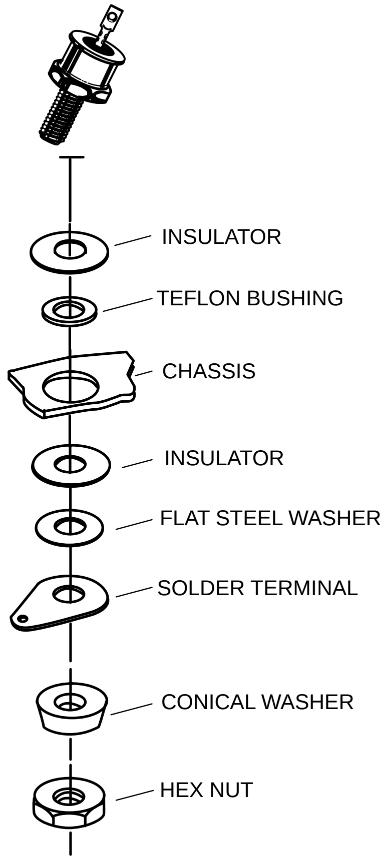
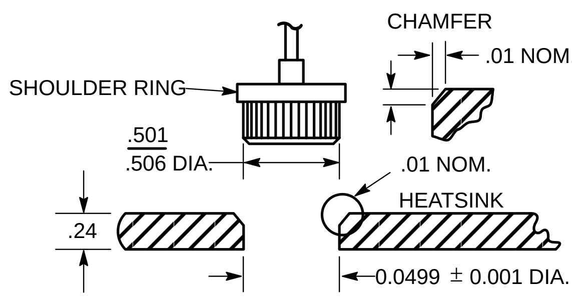


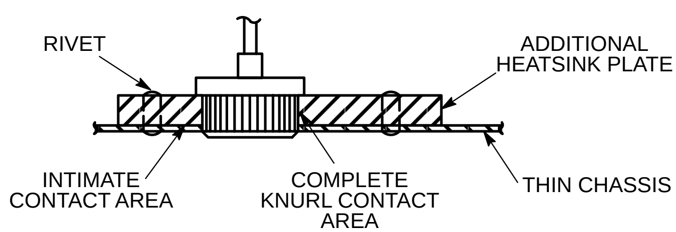
Figure 229. Isolating Hardware Used for a Nonisolated Stud–Mount Package

Tab Mount

The tab mount class is also used for rectifier diodes as illustrated in Figure 233. Mounting considerations are similar to that for the popular TO–220 package, whose suggested mounting arrangements and hardware are shown in Figure 234. The rectangular washer shown in the parts list under column a is used to minimize distortion of the mounting flange; excessive distortion could cause damage to the semiconductor chip. Use of the washer is only important when the size of the mounting hole exceeds 0.140 in. (6/32 clearance). Larger holes are needed to accommodate the lower insulating bushing when the screw is electrically connected to the case; however, the holes should not be larger than necessary to provide hardware clearance and should never exceed a diameter of 0.250 in. Flange distortion is also possible if excessive torque is used during mounting. A maximum torque of 8 in.–lb is suggested when using a 6/32 screw.



HeatSink Mounting



Thin–Chassis Mounting

The hole edge must be chamfered as shown to prevent shearing off the knurled edge of the case during press–in. The pressing force should be applied evenly on the shoulder ring to avoid tilting or canting of the case in the hold during the pressing operation. Also, the use of a thermal joint compound will be of considerable aid. The pressing force will vary from 250 to 1000 pounds, depending upon the heatsink material. Recommended harnesses are: copper–less than 50 on the Rockwell F scale; Aluminum–less than 65 on the Brinell scale. A heatsink as thin as 1/8″ may be used, but the interface thermal resistance will increase in direct proportion to the contact area. A thin chassis requires the addition of a backup plate.

Figure 230. Press–Fit Package

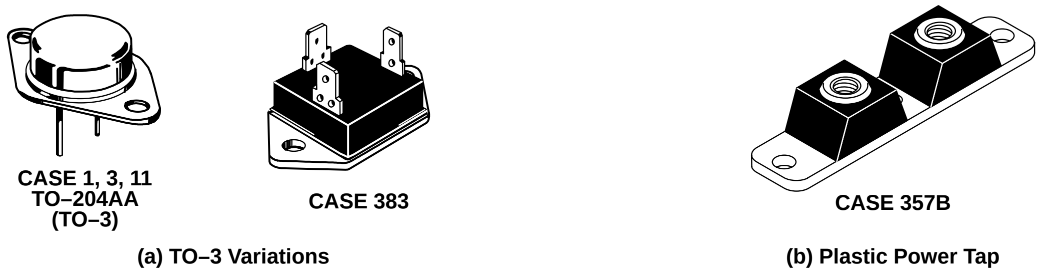


Figure 231. Flange–Mount Packages Used for Rectifiers

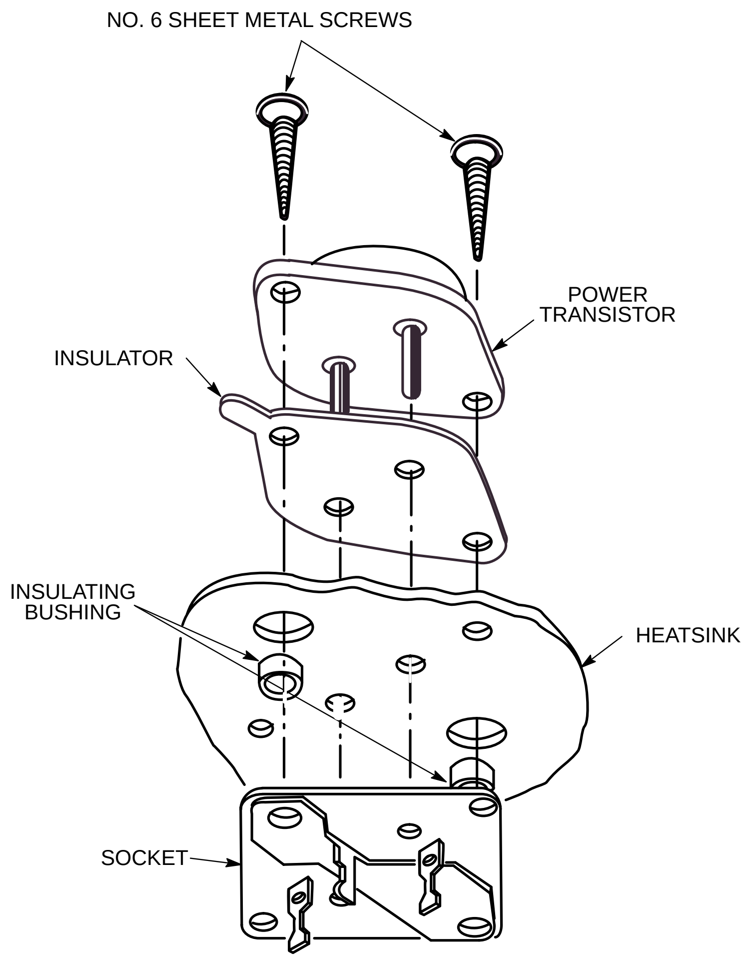


Figure 232. Hardware Used for a TO–204AA (TO–3) Flange–Mount Part

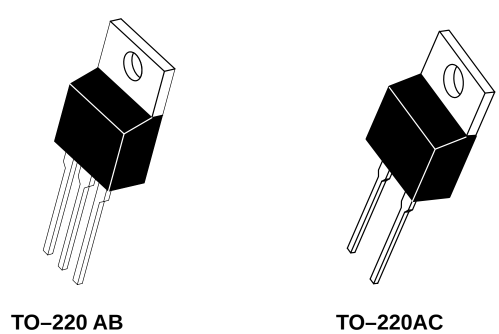


Figure 233. Several Types of Tab–Mount Parts

Care should be exercised to assure that the tool used to drive the mounting screw never comes in contact with the plastic body during the driving operation. Such contact can result in damage to the plastic body and internal device connections. To minimize this problem, some TO–220 packages have a chamfer on one end. Packages without a chamfer require a spacer or combination spacer and isolation bushing to raise the screw head above the top surface of the plastic.

The popular TO–220 package and others of similar construction, lift off the mounting surface as pressure is applied to one end. (See Appendix A, Figure 236) To counter this tendency, at least one hardware manufacturer [4] offers a hard plastic cantilever beam that applies more even pressure on the tab. In addition, it separates the mounting screw from the metal tab. Tab mount parts may also be effectively mounted with clips as shown in Figure 235. To obtain high pressure without cracking the case, a pressure spreader bar should be used under the clip. Interface thermal resistance with the cantilever beam or clips can be lower than with screw mounting.

In situations where a tab mount package is making direct contact with the heatsink, an eyelet may be used; provided sharp blows or impact shock is avoided.

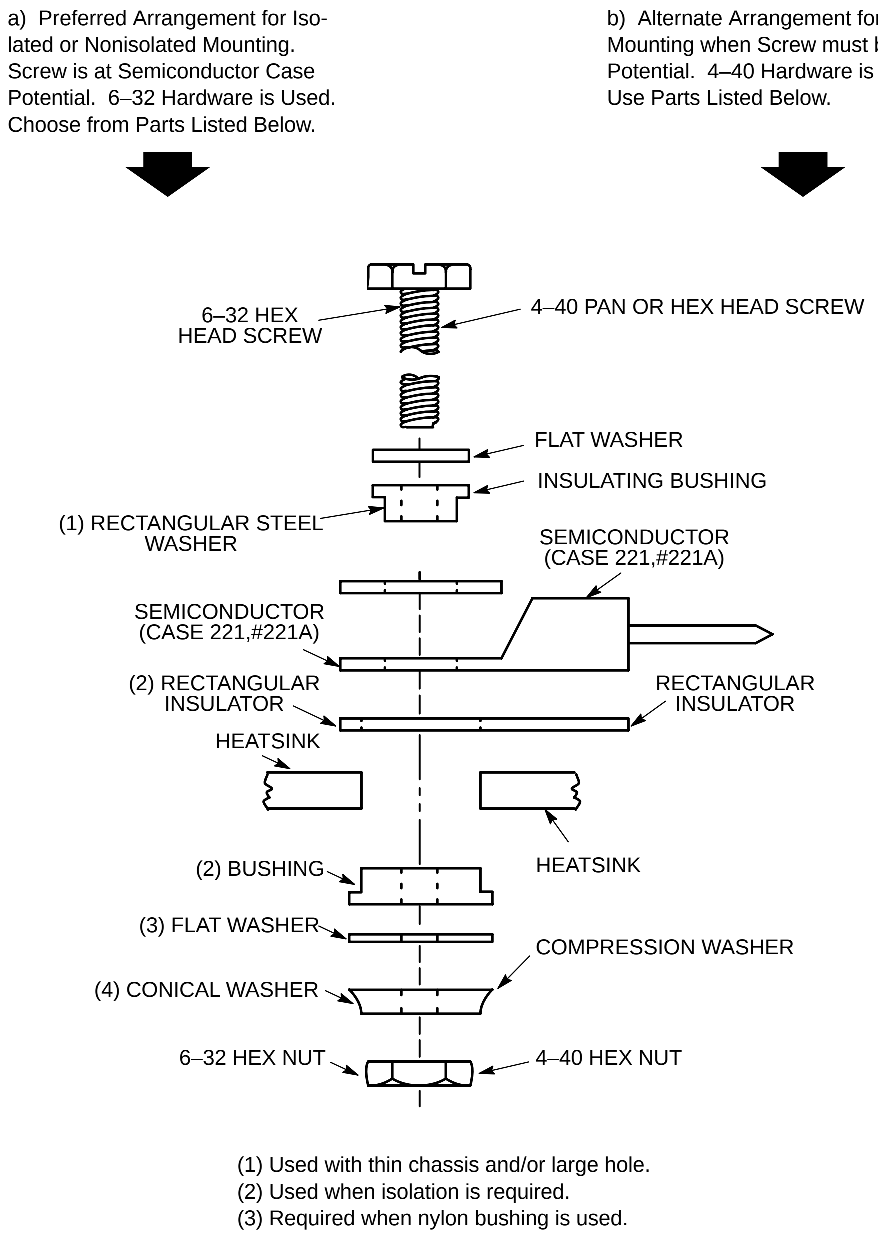


Figure 234. Mounting Arrangements for Tab Mount TO–220 Package

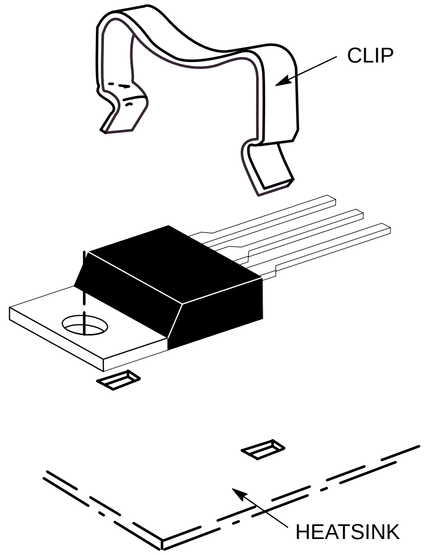


Figure 235. Clip Mounting a TO–220 Mount Package

References

1. Catalog #87–HS–9, page 8; Thermalloy, Inc., P.O. Box 810839, Dallas, Texas 75381–0839.
2. ITW Shakeproof, St. Charles Road, Elgin, IL 60120.
3. Robert Baston, Elliot Fraunglass and James P. Moran, “Heat Dissipation Through Thermally Conductive Adhesives,” EMTAS ‘83 Conference, Feb. 1–3, Phoenix, AZ.
4. Catalog, Edition 18, Richco Plastic Co., 5825 N. Tripp Ave., Chicago, IL 60546.

Appendix A

Measurement of Interface Thermal Resistance

Measurement of Interface Thermal Resistance

Measuring the interface thermal resistance $R_{\theta CS}$ appears deceptively simple. All that's apparently needed is a thermocouple on the semiconductor case, a thermocouple on the heatsink, and a means of applying and measuring DC power. However, $R_{\theta CS}$ is proportional to the amount of contact area between the surfaces and consequently is affected by surface flatness and finish and the amount of pressure on the surfaces. The fastening method may also be a factor. In addition, placement of the thermocouples can have a significant influence upon the results. Consequently, values for interface thermal resistance presented by different manufacturers are not in good agreement. Fastening methods and thermocouple locations are considered in this Appendix.

When fastening the test package in place with screws, thermal conduction may take place through the screws, for example, from the flange ear on a TO–3 package directly to the heatsink. This shunt path yields values which are artificially low for the insulation material and dependent upon screw head contact area and screw material. MIL–I–49456 allows screws to be used in tests for interface thermal resistance probably because it can be argued that this is "application oriented."

Thermalloy takes pains to insulate all possible shunt conduction paths in order to more accurately evaluate insulation materials. The ON Semiconductor fixture uses an insulated clamp arrangement to secure the package which also does not provide a conduction path.

As described previously, some packages, such as a TO–220, may be mounted with either a screw through the tab or a clip bearing on the plastic body. These two methods often yield different values for interface thermal resistance. Another discrepancy can occur if the top of the package is exposed to the ambient air where radiation and convection can take place. To avoid this, the package should be covered with insulating foam. It has been estimated that a 15 to 20% error in $R_{\theta CS}$ can be incurred from this source.

Another significant cause for measurement discrepancies is the placement of the thermocouple to measure the semiconductor case temperature. Consider the TO–220 package shown in Figure 236. The mounting pressure at one end causes the other end – where the die is located – to lift off the mounting surface slightly. To improve contact, ON Semiconductor TO–220 Packages are slightly concave. Use of a spreader bar under the screw lessens the lifting, but some is inevitable with a package of this structure. Three thermocouple locations are shown:

1. The ON Semiconductor location is directly under the die reached through a hole in the heatsink. The thermocouple is held in place by a spring which forces the thermocouple into intimate contact with the bottom of the semi's case.

1. The JEDEC location is close to the die on the top surface of the package base reached through a blind hole drilled through the molded body. The thermocouple is swaged in place.
2. The Thermalloy location is on the top portion of the tab between the molded body and the mounting screw. The thermocouple is soldered into position.

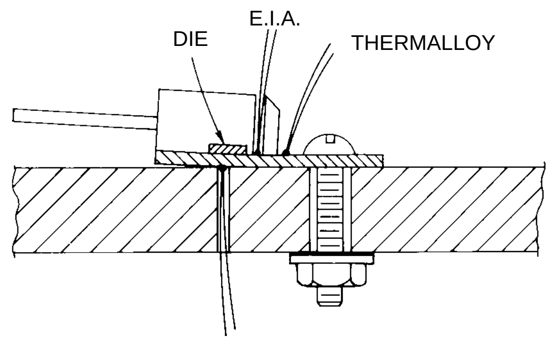


Figure 236. JEDEC TO–220 Package Mounted to Heatsink Showing Various Thermocouple Locations and Lifting Caused by Pressure at One End

Temperatures at the three locations are generally not the same. Consider the situation depicted in the figure. Because the only area of direct contact is around the mounting screw, nearly all the heat travels horizontally along the tab from the die to the contact area. Consequently, the temperature at the JEDEC location is hotter than at the Thermalloy location and the ON Semiconductor location is even hotter. Since junction–to–sink thermal resistance must be constant for a given test setup, the calculated junction–to–case thermal resistance values decrease and case–to–sink values increase as the "case" temperature thermocouple readings become warmer. Thus, the choice of reference point for the "case" temperature is quite important.

There are examples where the relationship between the thermocouple temperatures are different from the previous situation. If a mica washer with grease is installed between the semiconductor package and the heatsink, tightening the screw will not bow the package; instead, the mica will be deformed. The primary heat conduction path is from the die through the mica to the heatsink. In this case, a small temperature drop will exist across the vertical dimension of the package mounting base so that the thermocouple at the JEDEC location will be the hottest. The thermocouple temperature at the Thermalloy location will be lower, but close to the temperature at the JEDEC location as the lateral heat flow is generally small. The ON Semiconductor location will be coolest.

The JEDEC location is chosen to obtain the highest temperature on the case. It is of significance because power ratings are supposed to be based on this reference point. Unfortunately, the placement of the thermocouple is tedious and leaves the semiconductor in a condition unfit for sale.

The ON Semiconductor location is chosen to obtain the highest temperature of the case at a point where, hopefully, the case is making contact to the heatsink. Once the special heatsink, to accommodate the thermocouple has been fabricated, this method lends itself to production testing and does not mark the device. However, this location is not easily accessible to the user.

The Thermalloy location is convenient and is often chosen by equipment manufacturers. However, it also blemishes the case and may yield results differing up to 1°C/W for a TO–220 package mounted to a heatsink without thermal grease and no insulator. This error is small when compared to the thermal resistance of heat dissipaters often used with this package, since power dissipation is usually a few watts. When compared to the specified junction–to–case values of some of the higher power semiconductors becoming available, however, the difference becomes significant and it is important that the semiconductor manufacturer and equipment manufacturer use the same reference point.

Another JEDEC method of establishing reference temperatures utilizes a soft copper washer (thermal grease is used) between the semiconductor package and the heatsink. The washer is flat to within 1 mil/inch, has a finish better than 63 μ–inch, and has an imbedded thermocouple near its center. This reference includes the interface resistance under nearly ideal conditions and is, therefore, application–oriented. It is also easy to use, but has not become widely accepted.

A good way to improve confidence in the choice of case reference point is to also test for junction–to–case thermal resistance while testing for interface thermal resistance. If the junction–to–case values remain relatively constant as insulators are changed, torque varied, etc., then the case reference point is satisfactory.

Appendix B

NCP1200 10 W AC/DC Power Supply

NCP1200

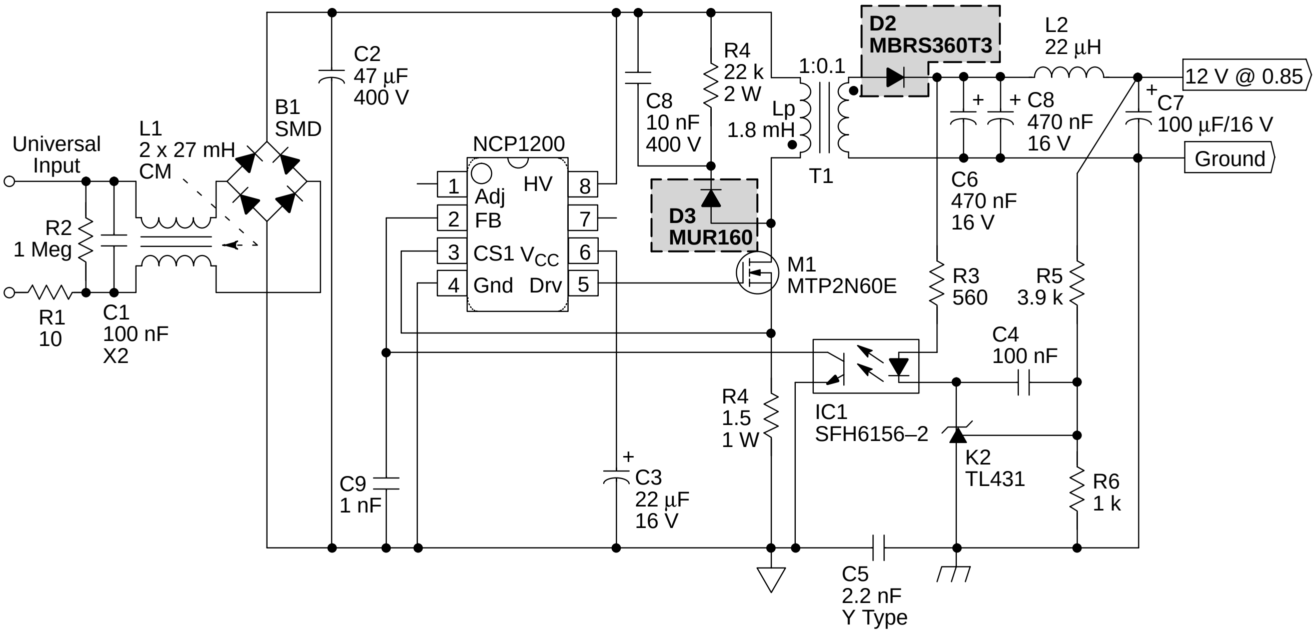


Figure 237. A complete 10 W AC/DC wall adapter featuring an EMI filter.

Additional Information

The following information is available on our web site at **www.onsemi.com** or from the ON Semiconductor Literature Distribution Center at 1–800–344–3860 or 1–303–675–2175 in hard copy or CDROM

1. NCP1200/D, Technical Data Sheet, "PWM Current–Mode Controller for Low–Power Universal Off–Line Supplies."
2. AND8032/D, Application Note, "Conducted EMI Filter Design for the NCP1200.", Christophe Basso.
3. NCP1200PAK/D, "NCP1200 Literature Pack."
4. MBRS320T3/D, Technical Data Sheet, "Surface Mount Schottky Power Rectifier."
5. MUR120/D, Technical Data Sheet, " MUR120 Series, SWITCHMODE™ Power Rectifiers."

Index

Alphabetical Subject Index

Alphabetical Subject Index

A

B

C

C (continued)

D

E

P

R

R (continued)

S

T

U

V

W

ON SEMICONDUCTOR MAJOR WORLDWIDE SALES OFFICES AND REPRESENTATIVES

UNITED STATES

ALABAMA
Huntsville 256–774–1000

CALIFORNIA
Encino 818–654–9040
Irvine 949–623–8485
Sacramento (Sales Rep) 916–652–0268
San Diego 858–621–5100
Santa Clara (Sales Rep) 408–350–4800

COLORADO
Littleton 303–256–5884

FLORIDA
Boca Raton 561–995–1466
Tampa 813–286–6181

GEORGIA
Duluth 888–793–3435

ILLINOIS
Chicago 847–330–6959

INDIANA
Kokomo 765–865–2085
Carmel (Sales Rep) 317–848–9958

MASSACHUSETTS
Boston 781–376–4223

MICHIGAN
Livonia 734–953–6704

MINNESOTA
Plymouth 763–249–2360

NORTH CAROLINA
Raleigh 919–785–5025

PENNSYLVANIA
Philadelphia/Horsham 215–997–4340

TEXAS
Dallas (Sales Rep) 972–680–2800

UNITED STATES (continued)

UTAH
Draper (Sales Rep) 801–572–4010

WASHINGTON
Seattle 425–646–7374

WASHINGTON, D.C.
. 410–884–4086

CANADA

QUEBEC
Pointe Claire 514–695–4599

ONTARIO
Toronto 905–812–0092

INTERNATIONAL

BRAZIL
Sao Paulo 55–011–3030–5244

CHINA
Beijing 86–10–8518–2323
Chengdu 86–28–678–4078
Shenzhen 86–755–209–1128
Shanghai 86–21–6875–6677

CZECH REPUBLIC
Roznov 420 (0) 651–667–141

FINLAND
Vantaa 358 (0) 9–85–666–460

FRANCE
Paris 33 (0) 1–39–26–41–00

GERMANY
Munich 49 (0) 89–93–0808–0

HONG KONG
Hong Kong 852–2689–0088

INTERNATIONAL (continued)

ISRAEL
Herzelia 972 (0) 9–9609–111

ITALY
Milan 39 (0) 2–530–0951

JAPAN
Tokyo 81–3–5740–2700

KOREA
Seoul 82–2–528–2700

MALAYSIA
Penang 60–4–226–9368

MEXICO
Guadalajara 523–669–9100

PHILIPPINES
Muntinlupa City 63–2–842–7141

PUERTO RICO
San Juan 787–641–4100

SINGAPORE
Singapore 65–298–1768

SWEDEN
Solna 46 (0) 8–5090–4680

TAIWAN
Taipei 886–2–2718–9961

THAILAND
Bangkok 66–2–653–5031

UNITED KINGDOM
Aylesbury 44 (0) 1296–610400

N. AMERICAN TECHNICAL SUPPORT
1–800–282–9855 Toll Free

ON SEMICONDUCTOR STANDARD DOCUMENT TYPE DEFINITIONS

DATA SHEET CLASSIFICATIONS

A Data Sheet is the fundamental publication for each individual product/device, or series of products/devices, containing detailed parametric information and any other key information needed in using, designing–in or purchasing of the product(s)/device(s) it describes. Below are the three classifications of Data Sheet: Product Preview; Advance Information; and Fully Released Technical Data

PRODUCT PREVIEW

A Product Preview is a summary document for a product/device under consideration or in the early stages of development. The Product Preview exists only until an "Advance Information" document is published that replaces it. The Product Preview is often used as the first section or chapter in a corresponding reference manual. The Product Preview displays the following disclaimer at the bottom of the first page: "This document contains information on a product under development. ON Semiconductor reserves the right to change or discontinue this product without notice."

ADVANCE INFORMATION

The Advance Information document is for a device that is NOT fully qualified, but is in the final stages of the release process, and for which production is eminent. While the commitment has been made to produce the device, final characterization and qualification may not be complete. The Advance Information document is replaced with the "Fully Released Technical Data" document once the device/part becomes fully qualified. The Advance Information document displays the following disclaimer at the bottom of the first page: "This document contains information on a new product. Specifications and information herein are subject to change without notice."

FULLY RELEASED TECHNICAL DATA

The Fully Released Technical Data document is for a product/device that is in full production (i.e., fully released). It replaces the Advance Information document and represents a part that is fully qualified. The Fully Released Technical Data document is virtually the same document as the Product Preview and the Advance Information document with the exception that it provides information that is unavailable for a product in the early phases of development, such as complete parametric characterization data. The Fully Released Technical Data document is also a more comprehensive document than either of its earlier incarnations. This document displays no disclaimer, and while it may be informally referred to as a "data sheet," it is not labeled as such.

DATA BOOK

A Data Book is a publication that contains primarily a collection of Data Sheets, general family and/or parametric information, Application Notes and any other information needed as reference or support material for the Data Sheets. It may also contain cross reference or selector guide information, detailed quality and reliability information, packaging and case outline information, etc.

APPLICATION NOTE

An Application Note is a document that contains real–world application information about how a specific ON Semiconductor device/product is used, or information that is pertinent to its use. It is designed to address a particular technical issue. Parts and/or software must already exist and be available.

SELECTOR GUIDE

A Selector Guide is a document published, generally at set intervals, that contains key line–item, device–specific information for particular products or families. The Selector Guide is designed to be a quick reference tool that will assist a customer in determining the availability of a particular device, along with its key parameters and available packaging options. In essence, it allows a customer to quickly "select" a device. For detailed design and parametric information, the customer would then refer to the device's Data Sheet. The *Master Components Selector Guide* (SG388/D) is a listing of **ALL** currently available ON Semiconductor devices.

REFERENCE MANUAL

A Reference Manual is a publication that contains a comprehensive system or device–specific descriptions of the structure and function (operation) of a particular part/system; used overwhelmingly to describe the functionality or application of a device, series of devices or device category. Procedural information in a Reference Manual is limited to less than 40 percent (usually much less).

HANDBOOK

A Handbook is a publication that contains a collection of information on almost any give subject which does not fall into the Reference Manual definition. The subject matter can consist of information ranging from a device specific design information, to system design, to quality and reliability information.

ADDENDUM

A documentation Addendum is a supplemental publication that contains missing information or replaces preliminary information in the primary publication it supports. Individual addendum items are published cumulatively. The Addendum is destroyed upon the next revision of the primary document.

ON Semiconductor and ON are trademarks of Semiconductor Components Industries, LLC (SCILLC). SCILLC reserves the right to make changes without further notice to any products herein. SCILLC makes no warranty, representation or guarantee regarding the suitability of its products for any particular purpose, nor does SCILLC assume any liability arising out of the application or use of any product or circuit, and specifically disclaims any and all liability, including without limitation special, consequential or incidental damages. "Typical" parameters which may be provided in SCILLC data sheets and/or specifications can and do vary in different applications and actual performance may vary over time. All operating parameters, including "Typicals" must be validated for each customer application by customer's technical experts. SCILLC does not convey any license under its patent rights nor the rights of others. SCILLC products are not designed, intended, or authorized for use as components in systems intended for surgical implant into the body, or other applications intended to support or sustain life, or for any other application in which the failure of the SCILLC product could create a situation where personal injury or death may occur. Should Buyer purchase or use SCILLC products for any such unintended or unauthorized application, Buyer shall indemnify and hold SCILLC and its officers, employees, subsidiaries, affiliates, and distributors harmless against all claims, costs, damages, and expenses, and reasonable attorney fees arising out of, directly or indirectly, any claim of personal injury or death associated with such unintended or unauthorized use, even if such claim alleges that SCILLC was negligent regarding the design or manufacture of the part. SCILLC is an Equal Opportunity/Affirmative Action Employer.

PUBLICATION ORDERING INFORMATION

Literature Fulfillment:
Literature Distribution Center for ON Semiconductor
P.O. Box 5163, Denver, Colorado 80217 USA
Phone: 303-675-2175 or 800-344-3860 Toll Free USA/Canada
Fax: 303-675-2176 or 800-344-3867 Toll Free USA/Canada
Email: ONlit@hibbertco.com

N. American Technical Support: 800-282-9855 Toll Free USA/Canada

JAPAN: ON Semiconductor, Japan Customer Focus Center
4-32-1 Nishi-Gotanda, Shinagawa-ku, Tokyo, Japan 141-0031
Phone: 81-3-5740-2700
Email: r14525@onsemi.com

ON Semiconductor Website: http://onsemi.com

For additional information, please contact your local Sales Representative